

Gane Samb LO

Weak Convergence (IA)

-

Sequences of Random Vectors

*Statistics and Probability African Society
(SPAS) Books Series.
Calgary, Alberta. 2016.*

DOI : <http://dx.doi.org/10.16929/sbs/2016.0001>

ISBN 978-2-9559183-1-9

SPAS Series Books

Advisers

List of published books

Library of Congress Cataloging-in-Publication Data

Gane Samb LO, 1958-

Weak Convergence (IA). Sequences of Random Vectors.

SPAS Books Series, 2016.

Copyright ©Statistics and Probability African Society (SPAS).

DOI : 10.16929/sbs/2016.0001

ISBN 978-2-9559183-1-9

Author : Gane Samb LO

Emails:

gane-samb.lo@ugb.edu.sn, ganesamblo@ganesamblo.net.

Url's:

www.ganesamblo@ganesamblo.net

www.statpas.net/cva.php?email.ganesamblo@yahoo.com.

Affiliations.

Main affiliation : University Gaston Berger, UGB, SENEGAL.

African University of Sciences and Technology, AUST, ABuja, Nigeria.

Affiliated as a researcher to : LSTA, Pierre et Marie Curie University, Paris VI, France.

Teaches or has taught at the graduate level in the following universities:

Saint-Louis, Senegal (UGB)

Banjul, Gambia (TUG)

Bamako, Mali (USTTB)

Ouagadougou - Burkina Faso (UJK)

African Institute of Mathematical Sciences, Mbour, SENEGAL, AIMS.

Franceville, Gabon

Dedicaces.

To my wife Mbaye Ndaw Fall who is accompanying for decades with love and patience

Acknowledgement of Funding.

The author acknowledges continuous support of the World Bank Excellence Center in Mathematics, Computer Sciences and Intelligence Technology, CEA-MITIC. His research projects in 2014, 2015 and 2016 are funded by the University of Gaston Berger in different forms and by CEA-MITIC.

Weak Convergence (IA). Sequences of Random Vectors

ABSTRACT. (English) This monograph aims at presenting the core weak convergence theory for sequences of random vectors with values in \mathbb{R}^k . In some places, a more general formulation in metric spaces is provided. It lays out the necessary foundation that paves the way to applications in particular subfields of the theory. In particular, the needs of Asymptotic Statistics are addressed. A whole chapter is devoted to weak convergence in \mathbb{R} where specific tools, for example for handling weak convergence of sequences using independent and identically distributed random variables such that the Renyi's representations by means of standard uniform or exponential random variables, are stated. The functional empirical process is presented as a powerful tool for solving a considerable number of asymptotic problems in Statistics. The text is written in a self-contained approach whith the proofs of all used results at the exception of the general Skorohod-Wichura Theorem.

(Français) Cet ouvrage a l'ambition de présenter le noyau dur de la théorie de la convergence vague de suite de vecteurs aléatoires dans \mathbb{R}^k . Autant que possible, dans certaines situations, la théorie générale dans des espaces métriques est donnée. Il prépare la voie à une spécialisation dans certains sous-domaines de la convergence vague. En particulier, les besoins de la statistique asymptotique ont été satisfaits. Un chapitre de l'ouvrage concerne la convergence vague dans \mathbb{R} avec des outils spécifiques, par exemple, pour étudier les suites de variables aléatoires indépendantes et identiquement distribuées tels que la représentation de Renyi au moyen de variables aléatoires uniformes ou exponentielles standard. Le processus empirique fonctionnel est introduit comme un outil puissant pour étudier des problèmes asymptotiques en Statistiques. Le texte est rédigé dans une approche auto-citante avec toutes les preuves des résultats utilisés, à l'exception du Théorème de Skorohod-Wichura.

Keywords. Weak convergence; Convergence in distribution; Portmanteau Theorem; Probability Laws characterization; Distribution functions; Characteristic functions; Probability density functions; Random Walks; Empirical processes; Multinomial Laws; Relative compactness; Asymptotic and uniform tightness; Continuous mapping theorem; Renyi and Malquist representations; Order Statistics; Multivariate Delta methods; Functional empirical process.

AMS 2010 Classification Subjects : 60XXX; 62G30

Contents

General Preface	1
General Preface of Our Series of Weak Convergence	3
Preface of The Series Weak Convergence : Sequence of Random Vectors	5
Chapter 1. Review of Usual Weak Convergence Results in \mathbb{R}^k	9
1. Introduction	9
2. Weak Convergence in \mathbb{R}^k	9
3. Examples of Weak Convergence in \mathbb{R}	13
4. Examples of Convergence in \mathbb{R}^k	24
5. Invariance principle	34
Chapter 2. Weak Convergence Theory	37
1. Introduction	37
2. Definition, Unicity and Portmanteau Theorem	37
3. Continuous Mapping Theorem	45
4. Space \mathbb{R}^k	47
5. Theorem of Scheffé	58
6. Weak Convergence and Convergence in Probability on one Probability Space	61
7. Appendix	68
Chapter 3. Uniform Tightness and Asymptotic Tightness	79
1. Introduction	79
2. Tightness	86
3. Compacity Theorem for weak convergence in \mathbb{R}^k	90
4. Applications	95
Chapter 4. Specific Tools for Weak Convergence in \mathbb{R}	99
1. Generalized inverses of monotone functions	99
2. Applications of Generalized functions	110
3. Representation of Renyi for iid sequences of random Variables	111

Chapter 5. The functional Empirical Process As a General Tools in Asymptotic Statistics	121
1. Using the small o 's and the big O 's	121
2. Stochastic o 's and O 's	121
3. Delta Methods	139
4. Using the Functional Empirical Process in Asymptotic Statistics	144
Chapter 6. Elements of Theory of Functions and Real Analysis	155
1. Review on limits in $\overline{\mathbb{R}}$. What should not be ignored on limits.	155
2. Miscellaneous facts	167
Bibliography	169

General Preface

This textbook is the first of series whose ambition is to cover broad part of Probability Theory and Statistics. These textbooks are intended to help learners and readers, of all levels, to train themselves.

As well, they may constitute helpful documents for professors and teachers for both courses and exercises. For more ambitious people, they are only starting points towards more advanced and personalized books. So, these textbooks are kindly put at the disposal of professors and learners.

Our textbooks are classified into categories.

A series of introductory books for beginners. Books of this series are usually accessible to student of first year in universities. They do not require advanced mathematics. Books on elementary probability theory and descriptive statistics are to be put in that category. Books of that kind are usually introductions to more advanced and mathematical versions of the same theory. The first prepare the applications of the second.

A series of books oriented to applications. Students or researchers in very related disciplines such as Health studies, Hydrology, Finance, Economics, etc. may be in need of Probability Theory or Statistics. They are not interested by these disciplines by themselves. Rather, the need to apply their findings as tools to solve their specific problems. So adapted books on Probability Theory and Statistics may be composed to on the applications of such fields. A perfect example concerns the need of mathematical statistics for economists who do not necessarily have a good background in Measure Theory.

A series of specialized books on Probability theory and Statistics of high level. This series begin with a book on Measure Theory, its counterpart of probability theory, and an introductory book on topology. On that basis, we will have, as much as possible, a coherent

presentation of branches of Probability theory and Statistics. We will try to have a self-contained, as much as possible, so that anything we need will be in the series.

Finally, **research monographs** close this architecture. The architecture should be so large and deep that the readers of monographs booklets will find all needed theories and inputs in it.

We conclude by saying that, with only an undergraduate level, the reader will open the door of anything in Probability theory and statistics with **Measure Theory and integration**. Once this course validated, eventually combined with two solid courses on topology and functional analysis, he will have all the means to get specialized in any branch in these disciplines.

Our collaborators and former students are invited to make live this trend and to develop it so that the center of Saint-Louis becomes or continues to be a reknown mathematical school, especially in Probability Theory and Statistics.

General Preface of Our Series of Weak Convergence

The series Weak convergence is an open project with three categories.

The special series Weak convergence I consists of texts devoted to the core theory of weak convergence, each of them concentrated on the handling of one specific class of objects. The texts will have labels A , B , etc. Here are some examples.

- (1) Weak convergence of Random Vectors (IA).
- (2) Weak convergence of stochastic processes and empirical processes (IB).
- (3) Weak convergence of random measures (IC).
- (4) Weak convergence of fuzzy random measures (IC).

The special series Weak convergence II consists of textbooks related to the theory of weak convergence, each of them concentrated on one specialized field using weak convergence. Usually, these subfields are treated apart in the literature. Here, we want to put them in our general frame as continuations of the Weak Convergence Series I. Some examples are the following.

- (1) Weak laws of sums of independent random variables.
- (2) Weak laws of sums of associated random variables.
- (3) Univariate Extreme value Theory.

(4) Multivariate Extreme value Theory.

(5) Etc.

The special series Weak convergence III consists of textbooks focusing on statistical applications of Parts of the Weak Convergence Series I and Weak Convergence Series II. Examples :

(1) A handbook of Gaussian Asymptotic Distribution Using the Functional Empirical Process.

(2) A handbook of Statistical Estimation of the Extreme Value index.

(1) etc.

Preface of The Series Weak Convergence : Sequence of Random Vectors

The series Weak convergence (IA) concerns the theory of weak convergence of sequences of random vectors. Due to the theorem of Kolmogorov, stating broadly that the probability law of any random element is characterized by its finite distribution under the appropriate state spaces, the place of the distributions of random vectors is surely central to Probability Theory.

This motivated us to begin this series by the weak convergence of random vectors as the foundation of all the structure.

Another reason is that the needs of Asymptotic Statistics, which is one of the main motivations of the development of Weak Convergence Theory, generally does not need more than that. This booklet then gives to some readers exactly what they specifically.

This textbook focuses on the study of random elements in \mathbb{R}^k , $k \geq 1$. So the properties and the topology of \mathbb{R}^k are used.

But when only the general properties of the metric are used, we prefer to give the results in the general case where the studied sequences have their values in a metric space with a metric d .

The concept of tightness is essential in weak convergence theory. In this text, the Helly-Bray method is exclusively used.

This textbook is concluded by a chapter of the functional empirical process. Here, only the weak limits of its finite distributions are treated. We show how to use it for deriving asymptotic results in many research problems. With such tools, even at this somewhat elementary level of weak convergence, it is possible for readers to provide contributions in many research fields in Statistics and in applied related fields.

I wish you a pleasant reading and hope receiving your feedback.

To my wife Mbaye Ndaw Fall who is accompanying me since decades.

Saint-Louis, Calgary, Abuja, Bamako, Ouagadougou, 2016.

Preliminary Remarks and Notations.

WARNINGS

(1) In all this book, any unsepecified limit in presence with the subscripts n are meant as $n \rightarrow +\infty$.

(2) This textbook deals with general distribution functions F on \mathbb{R}^k , $k \geq 1$. The Lebesgues-Stieljes measure induced by a general distribution function is not necessarily a probability measure. If this induced Lebesgues-Stieljes is a probability measure, we precise this distribution function as a **probability distribution function**. As well, a distribution function of a random vector X of Lebesgues-Stieljes is implicately a probability distribution function although we do not say : the probability distribution function of X .

CHAPTER 1

Review of Usual Weak Convergence Results in \mathbb{R}^k

1. Introduction

In this chapter, we will see that most of the readers, actually know a considerable number of weak convergence results, even if they did not use this concept. What has to be done, on top of this review, is to present these individual results in the frame of a unified theory in the most general setting. This is the target of this book which will be given in the subsequent chapters.

Here, we are going to recall classical convergence results that any student should have encountered from the first courses in probability theory or in Statistics.

We begin to set the general frame of weak convergence in \mathbb{R}^k , $k \geq 1$. We will admit the statements in the following section. We will be able to establish their validity in Chapter 2, in particular in Theorem 4 of that chapter.

2. Weak Convergence in \mathbb{R}^k

Let us remind that the probability law of any vectorial random variable $X : (\Omega, \mathbb{A}, \mathbb{P}) \mapsto \mathbb{R}^k$ is characterized by

(a) its distribution function:

$$\mathbb{R}^k \ni x \mapsto F_X(x) = \mathbb{P}(X \leq x),$$

(b) its characteristic function (Here, i is the complex number such that $i^2 = -1$ with positive sinus, and $\langle ., . \rangle$ stands for the classical product space on \mathbb{R}^k)

$$\mathbb{R}^k \ni u \mapsto \Phi(u) = E(\exp(i \langle u, X \rangle)),$$

(c) its moment generating function (if it exists in a neighborhood of the null vector)

$$\mathbb{R}^k \ni u \mapsto \Psi_X(u) = E(\exp(\langle u, X \rangle)).$$

and

(d) its Radon-Nikodym derivative, or probability density distributive function (*pdf*), (if it exists), with respect (*w.r.t*) to a measure ν sur \mathbb{R}^k :

$$d\mathbb{P}/d\nu = f_X.$$

It is interesting that these characteristics also play the main roles in weak convergence through Theorem 4 we will prove in Chapter 2.

We have :

THEOREM 1. (**THEOREM-DEFINITION-LEMMA**) Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ be a sequence of random vectors $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ a random vector. Then the assertions (a) and (b) below are equivalent.

(a) For any $u \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \rightarrow \Phi_X(u) \text{ as } n \rightarrow +\infty.$$

(b) For any continuity point $u \in \mathbb{R}^k$ of F_X ,

$$F_{X_n}(x) \rightarrow F_X(x) \text{ as } n \rightarrow +\infty.$$

If one the assertions (a) or (b) holds, we say that the sequence X_n weakly converges to X , or X_n converges in distributions to X or X_n converges in law to X , as $n \rightarrow +\infty$, and we denote this by

$$X_n \rightsquigarrow X \text{ or } X_n \longrightarrow_d X \text{ or } X_n \xrightarrow{\mathcal{L}} X \text{ or } X_n \xrightarrow{w} X \text{ or } X_n \longrightarrow_w X$$

The weak limit is unique in distribution, meaning that if X_n weakly converges to X and to Y , then X and Y have the same distribution, that is $F_X = F_Y$ in the context of \mathbb{R}^k .

We also have the following sufficiency weak convergence conditions.

(c) If the moment functions Ψ_{X_n} exist on B_n , $n \geq 1$ and Ψ_X exist on B , where the B_n and B are neighborhood of 0, and if for any $x \in B$,

$$\Psi_{X_n}(x) \rightarrow \Psi_X(x) \text{ as } n \rightarrow +\infty,$$

then X_n weakly converges to X .

(d) Finally, suppose that the probability distribution $\mathbb{P}_n(\circ) = \mathbb{P}_n(X_n \in \circ)$, $n \geq 1$, and $\mathbb{P}_X(\circ) = \mathbb{P}_\infty(X \in \circ)$ have Radon-Nikodym derivatives with respect to a measure ν on \mathbb{R}^k , denoted by

$$d\mathbb{P}_n/d\nu = f_{X_n} \quad n \geq 1, \quad d\mathbb{P}_X/d\nu = f_X.$$

If for any $x \in D_X = \{x, f_X(x) > 0\}$,

$$f_{X_n}(x) \rightarrow f_X(x) \text{ as } n \rightarrow +\infty,$$

then $X_n \rightsquigarrow X$.

We have the following last point.

(e) Assume that the sequence $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ weakly converges to $X \in \mathbb{R}^k$, as $n \rightarrow +\infty$ and let A be a real (m, k) -matrix with $m \geq 1$. Then $\{AX_n, n \geq 1\} \subset \mathbb{R}^m$ weakly converges to $AX \in \mathbb{R}^m$.

Remark. Point (e) of Theorem 1 above is a consequence of the continuous mapping Theorem 6 in Chapter 2.

In summary, the weak convergence in \mathbb{R}^k holds when the distribution functions, the characteristic functions, the moment functions (if they exist) or the probability density functions (if they exist) with respect to the same measure ν , pointwisely converge to the distribution function, or to the characteristic function or to moment function (if it exists), or to the probability density function (if it exists) with respect to ν of a probability measure in \mathbb{R}^k . In the case of pointwise convergence of the distribution functions, only matters the convergence for continuity points of the limiting distribution functions.

All this is awesome and gives us pretty well tools to deal with weak convergence. The examples given below form the core set of examples

you cannot ignore.

But before we proceed to this review, we need a handsome criterion derived from the convergence of characteristic functions.

PROPOSITION 1. (*Wold Criterion*). *The sequence $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ weakly converges to $X \in \mathbb{R}^k$, as $n \rightarrow +\infty$ if and only if for any $a \in \mathbb{R}^k$, the sequence $\{\langle a, X_n \rangle, n \geq 1\} \subset \mathbb{R}$ weakly converges to $\langle a, X \rangle \in \mathbb{R}$ as $n \rightarrow +\infty$.*

Proof. The proof is quick and uses the notation above. Suppose that X_n weakly converges to X in \mathbb{R}^k as $n \rightarrow +\infty$. By using the convergence of characteristic functions, we have for any $u \in \mathbb{R}^k$

$$\mathbb{E}(\exp(i \langle X_n, u \rangle)) \rightarrow \mathbb{E}(\exp(i \langle X, u \rangle)) \quad \text{as } n \rightarrow +\infty.$$

It follows that for any $a \in \mathbb{R}^k$ and for any $t \in \mathbb{R}$, we have

$$(2.1) \quad \mathbb{E}(\exp(it \langle X_n, a \rangle)) \rightarrow \mathbb{E}(\exp(it \langle X, a \rangle)) \quad \text{as } n \rightarrow +\infty,$$

that is, by taking $u = ta$ in the formula above, and by denoting $Z_n = \langle X_n, a \rangle$ and $Z = \langle X, a \rangle$, we have

$$\mathbb{E}(\exp(itZ_n)) \rightarrow \mathbb{E}(\exp(itZ)) \quad \text{as } n \rightarrow +\infty.$$

This means that $Z_n \rightsquigarrow Z$, that is $\langle a, X_n \rangle$ weakly converges to $\langle a, X \rangle$.

Conversely, suppose that for any $a \in \mathbb{R}^k$, the sequence $\{\langle a, X_n \rangle, n \geq 1\} \subset \mathbb{R}$ weakly converges to $\langle a, X \rangle \in \mathbb{R}$ as $n \rightarrow +\infty$. Then by taking $t = 1$ in (4.17) we get for any $a = u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i \langle X_n, u \rangle)) \rightarrow \mathbb{E}(\exp(i \langle X, u \rangle)) \quad \text{as } n \rightarrow +\infty.$$

which means that $X_n \rightsquigarrow X$ as $n \rightarrow +\infty$.

3. Examples of Weak Convergence in \mathbb{R}

3.1. Weak Convergence of a sequence of Hypergeometric random variables to a Binomial random variable. Let X_N be a random variable following a Hypergeometric law $\mathcal{H}(N, M, n)$ with $M/N \rightarrow p$, $N \rightarrow \infty$, n being fixed. Then X_N weakly converges to a Binomial random variable X , that is $X \sim \mathcal{B}(n, p)$.

Proof. Let us use the probability density functions with respect to the counting measure ν on \mathbb{N} . We have

$$f_{X_n}(k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, 0 \leq k \leq \min(n, M).$$

Suppose that $M/N \rightarrow p$, $N \rightarrow \infty$. We have

$$\begin{aligned} f_{X_n}(k) &= \frac{M!}{k!(M-k)!} \frac{(N-M)!}{(n-k)!(N-M-(n-k))!} \frac{n!(N-n)!}{N!} \\ &= \frac{n!}{k!(n-k)!} \times \frac{M!}{(M-k)!} \times \frac{(N-M)!}{(N-M-(n-k))!} \times \frac{(M-n)!}{N!} \\ &= \binom{n}{k} \times \left\{ \frac{M!}{(M-k)!} \right\} \left\{ \frac{(N-M)!}{(N-M-(n-k))!} \right\} \left\{ \frac{(M-n)!}{N!} \right\}. \end{aligned}$$

But

$$\begin{aligned} \left\{ \frac{M!}{(M-k)!} \right\} &= (M-k+1)(M-k+2)\dots(M-1)M \\ &= M^k \left(1 - \frac{k-1}{M}\right) \left(1 - \frac{k-2}{M}\right) \times \dots \times \left(1 - \frac{1}{M}\right) \\ &= M^k (1 + o(1)) \end{aligned}$$

since $M \rightarrow \infty$ and k is fixed. Next,

$$\begin{aligned} \left\{ \frac{(N-M)!}{(N-M-(n-k))!} \right\} &= (N-M-(n-k)+1) \times \dots \times (N-M-1)(N-M) \\ &= (N-M)^{n-k} \left(1 + \frac{n-k-1}{N-M}\right) \left(1 + \frac{n-k-2}{N-M}\right) \times \dots \times \left(1 + \frac{1}{N-M}\right) \\ &= (N-M)^{n-k} - 1 + o(1) \end{aligned}$$

since, also, $N - M = N(1 - M/N) \sim N(1 - p) \rightarrow \infty$ and $n - k$ is fixed. Finally

$$\begin{aligned} \left\{ \frac{(M - n)!}{N!} \right\} &= \frac{1}{(N - n + 1)(N - n + 2) \dots (N - 1)N} \\ &= \frac{1}{N^n (1 - \frac{n-1}{N})(1 - \frac{n-2}{N}) \dots (1 - \frac{1}{N})} \\ &= \frac{1}{N^n (1 + o(1))}. \end{aligned}$$

for similar reasons. In total for any $0 \leq k \leq n$

$$f_{X_n}(k) = \binom{n}{k} \left(\frac{M}{N} \right)^k \left(\frac{N - M}{N} \right)^{n-k} (1 + o(1)) \rightarrow \binom{n}{k} p^k (1 - p)^{n-k}.$$

Hence, for any k in the support set of the *pdf* of a $\mathcal{B}(n, p)$ random variable *w.r.t* to the counting measure ν , denoted

$$f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

we have

$$\forall (1 \leq k \leq n), f_{X_n}(k) \rightarrow f_X(k).$$

The proof is finished.

Useful remark in sampling technique theory. This result allows to treat drawing without replacement (which generates a hypergeometric law) may be approximated as a drawing with replacement (which gives a Binomial law) when the size of the global population is large. The idea behind this is the following : if we randomly draw a small number of individuals from a large set, it is almost improbable that we draw one individual more than one time.

3.2. Weak Convergence of a sequence of Binomial random variables to a Poisson random variable. Let X_n be a sequence of $\mathcal{B}(n, p)$ random variable with $p = p_n \rightarrow 0$ and $np_n \rightarrow \lambda$, $0 < \lambda$, as $n \rightarrow \infty$. Then X_n weakly converges to a Poisson random variable X with parameter λ , that is $X \sim \lambda$.

Proof. Let us use the moment generating functions. Let X_n be a sequence of $\mathcal{B}(n, p_n)$ random variable and X be a $\mathcal{P}(\lambda)$ random variable. We have

$$\Psi_{X_n}(t) = ((1 - p_n) + p_n e^t)^n \text{ for } n \geq 1; \Psi_X(t) = \exp(\lambda(e^t - 1)), t \in \mathbb{R}.$$

Put $\lambda_n = np_n \rightarrow \lambda$. For any fixed t , we have

$$\Psi_{X_n}(t) = \left(\frac{\lambda_n}{n} + \left(1 - \frac{\lambda_n}{n} \right) e^t \right)^n = \left(1 - \frac{\lambda_n(e^t - 1)}{n} \right)^n \rightarrow \exp(\lambda(e^t - 1)) = \Psi_X(t)$$

by the following classical results of Analysis :

$$\left(1 + \frac{x_n}{n} \right)^n \rightarrow e^x \text{ as } n \rightarrow +\infty \text{ whenever } x_n \rightarrow x \in \mathbb{R} \text{ as } n \rightarrow +\infty.$$

3.3. Weak Convergence of a sequence of Poisson random variable to a Gaussian random variable. Let Z_λ be a Poisson random variable with parameter $\lambda > 0$, that is $Z_\lambda \sim \mathcal{P}(\lambda)$. Then the random variable

$$\frac{Z_\lambda - \lambda}{\sqrt{\lambda}}$$

weakly converges to standard Gaussian random variable X , that is $X \sim \mathcal{N}(0, 1)$, as $\lambda \rightarrow +\infty$.

Proof. Let us use the moment generating functions. The moment generating function of $Z_\lambda \sim \mathcal{P}(\lambda)$ is

$$\Psi_{Z_\lambda}(t) = \exp(\lambda(e^t - 1)).$$

Set

$$Y(\lambda) = \frac{Z_\lambda}{\sqrt{\lambda}} = \frac{Z_\lambda - \mathbb{E}(X)}{\sigma_{Z_\lambda}}.$$

We have

$$\Psi_{Y(\lambda)}(u) = e^{-\sqrt{\lambda}} \times \varphi_Z(u/\sqrt{\lambda}) = e^{-\sqrt{\lambda}} \times \exp(\lambda(e^{u/\sqrt{\lambda}} - 1)).$$

As $\lambda \rightarrow \infty$, we may use the following expansion

$$\begin{aligned} \lambda(e^{u/\sqrt{\lambda}} - 1) &= \lambda\left(1 + \frac{u}{\sqrt{\lambda}} + \frac{u^2}{2\lambda} + O(\lambda^{-3/2}) - 1\right) \\ &= u\sqrt{\lambda} + \frac{u^2}{2} + O(\lambda^{-1/2}). \end{aligned}$$

Hence

$$\Psi_{Y(\lambda)}(u) = \exp\left(\frac{u^2}{2} + O(\lambda^{-1/2})\right) \rightarrow \exp(u^2/2).$$

We conclude that

$$\frac{Z_\lambda}{\sqrt{\lambda}} \rightarrow \mathcal{N}(0, 1)$$

as $\lambda \rightarrow \infty$.

3.4. Convergence of a sequence of Binomial random variables to a standard Gaussian random variable. Let X_n be a $\mathcal{B}(n, p)$ random variable with $p \in]0, 1[$ which is fixed. Then, as $n \rightarrow \infty$,

$$(3.1) \quad Z_n = \frac{X_n - np}{\sqrt{npq}} \rightsquigarrow \mathcal{N}(0, 1).$$

Proof. Let us use the moment generating functions. Let $X \sim \mathcal{B}(n, p)$. We have

$$\Psi_{X_n}(u) = (q + pe^u)^n.$$

where $q = 1 - p$. Then

$$(3.2) \quad \Psi_{(X_n - np)/\sqrt{npq}}(u) = e^{-\sqrt{np/q}} \times \Psi_{X_n}(u/\sqrt{npq}),$$

with

$$\Psi_X(u/\sqrt{npq}) = (q + pe^{u/\sqrt{npq}})^n.$$

The idea behind the coming computations is to use a second order expansion of $e^{u/\sqrt{npq}}$ in the neighborhood of 0 as $n \rightarrow \infty$ and u fixed. We get an expression of the form $1 + v_n$, where v_n tends to zero. Finally an expansion of the logarithm function $\log(1 + v_n)$ of order 2 is operated.

Hence, as $n \rightarrow \infty$ and u is fixed, we have,

$$e^{u/\sqrt{npq}} = 1 + \frac{u}{\sqrt{npq}} + \frac{u^2}{2npq} + O(n^{-3/2}).$$

Next,

$$(q + pe^{u/\sqrt{npq}}) = 1 + u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2}) = 1 + v_n$$

with

$$v_n = u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2}) \rightarrow 0.$$

Thus,

$$\begin{aligned} \log \left(1 + u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2}) \right) &= \log(1 + v_n) \\ &= v_n - \frac{1}{2}v_n^2 + O(v_n^3) \\ &= u\sqrt{p/nq} + \frac{u^2}{2nq} - \frac{pu^2}{2nq} + O(n^{-3/2}). \end{aligned}$$

Hence

$$\begin{aligned}
\Psi_{X_n}(u/\sqrt{npq}) &= (q + pe^{u/\sqrt{npq}})^n = \exp(n \log(q + pe^{u/\sqrt{npq}})) \\
&= \exp\left(n \left(u\sqrt{p/nq} + \frac{u^2}{2nq} - \frac{pu^2}{2nq} + O(n^{-3/2})\right)\right) \\
&= \exp\left(u\sqrt{np/q} + \frac{u^2}{2q} - \frac{pu^2}{2q} + O(n^{-1/2})\right) \\
&= e^{u\sqrt{np/q}} e^{u^2/2 + O(n^{-1/2})}.
\end{aligned}$$

By going back to (3.2), we arrive at

$$\Psi_{(X_n - np)/\sqrt{npq}}(u) \rightarrow \exp(u^2/2).$$

This is

$$(\beta(n, p) - np)/\sqrt{npq} \rightarrow_w \mathcal{N}(0, 1) \text{ as } n \rightarrow +\infty.$$

QED.

Remark. We will come back for a direct proof of this result using the central limit theorem stated just below.

3.5. Simple Central Limit Theorem in \mathbb{R} . The two last cases are special cases of a more general weak convergence theorem, called the central limit theorem (*CLT*) of Probability Theory. We say that a sequence of real random variables $(X_n)_{n \geq 1}$, for which each X_n has a positive finite second moment, satisfies the *CTL* property if and only if

$$\frac{X_n - E(X_n)}{\sigma_{X_n}}$$

weakly converges to Gaussian standard random variable. This, of course, is not always true. Here, we will see a simple case. Later, we will give a global solution of this problem in \mathbb{R} .

Let X_1, X_2, \dots be a sequence of real valued random variables which are independent and identically distributed (*iid*) random variables with common distribution function F with

$$E(X_i) = \mu = \int x dF(x) = 0, \sigma_{X_i}^2 = \sigma^2 = \int (x - \mu)^2 dF(x) = 1.$$

Put, for $n \geq 1$,

$$S_n = X_1 + \dots + X_n.$$

We have, as $n \rightarrow \infty$,

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1).$$

Proof. Consider the common characteristic functions

$$\mathbb{R} \ni u \mapsto \Phi_{X_i}(u) = E(e^{iuX_i}) = \Psi(u).$$

Since the second moment exists, we have the following expansion at order 2,

$$\begin{aligned} \Phi(u) &= 1 + iu\Phi'(0) + \frac{1}{2}u^2\Phi''(0) + O(u^3) \\ &= 1 - \frac{1}{2}u^2 + O(u^2) \end{aligned}$$

since

$$\Phi'(0) = i \mathbb{E}(X) = 0, \quad \Phi''(0) = -\mathbb{E}(X^2) = -1.$$

Thus

$$\Phi_{S_n/\sqrt{n}}(u) = (\Phi(u/\sqrt{n}))^n.$$

For u fixed, as $n \rightarrow \infty$,

$$\begin{aligned} \Phi_{S_n/\sqrt{n}}(u) &= (\Phi(u/\sqrt{n}))^n = \exp\left(n \log\left(1 - \frac{u^2}{2n} + O(n^{-3/2})\right)\right) \\ &= \exp\left(n\left(-\frac{u^2}{n} + O(n^{-3/2})\right)\right) \\ &= \exp(-u^2/2 + O(n^{-1/2})) \\ &\rightarrow \exp(-u^2/2). \end{aligned}$$

We just established

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1) \text{ as } n \rightarrow +\infty.$$

In a more general case of an *iid* sequence of random variables X_1, X_2, \dots with

$$\mathbb{E}(X_i) = \mu = \int x dF(x) = \mu, \sigma_{X_i}^2 = \sigma^2 = \int (x - \mu)^2 dF(x) = \sigma^2,$$

we apply the former result to the sequence $(X_i - \mu)/\sigma$, $i = 1, 2, \dots$ to get

$$\frac{1}{\sigma\sqrt{n}}(S_n - n\mu) \rightarrow \mathcal{N}(0, 1).$$

Let us give two examples of applications of the simple central limit theorem on the binomial trials.

Example 1 : Weak convergence of the binomial random variable.

We are going to prove the result (3.1) of Subsection 3.4 concerning the weak law of a sequence of binomial random variables as the number of trials, n , increases while the probability of success, $p \in]0, 1[$, is fixed. So we keep the notation of that subsection.

We know from the earlier courses on elementary Probability Theory we may find in a considerable number of books, especially in [6], with the current Probability Theory and Statistics Series, in Chapter 2, Lemma 1, that if $X_n \sim \mathcal{B}(n, p)$, then X_n is the sum of n independent Bernoulli $\mathcal{B}(p)$ random variables Y_1, \dots, Y_n such that

$$X_n = Y_1 + \dots + Y_n.$$

For each of the Y_i 's random variables, we have

$$\mathbb{E}(Y_i) = p \text{ and } \sigma^2 = \text{Var}(Y_i) = pq \text{ where } q = 1 - p.$$

Then, the random variable Z_n in Formula (3.1) becomes

$$Z_n = \frac{X_n - np}{\sqrt{npq}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)).$$

Hence, the weak convergence of Z_n to $\mathcal{N}(0, 1)$ as $n \rightarrow +\infty$, is a consequence of the simple central limit standard on \mathbb{R} .

Remark. This proof is quick and beautiful. The first proof is still useful. Because, we may be in a position to teach this result at a level where the central limit theorem is not available. Besides, this proof is part of History. In the same spirit, the oldest proof of this result goes back to 1732 by *de Moivre* and to 1801 by Laplace (see Loève [10], page 23). These historical methods can also be found in [6] and in [7] with a writing which is appropriate to beginners of first year of university.

Example 2 : Negative Binomial Law.

For a fixed integer $k \geq 1$, a Negative Binomial random variable X_k is defined relatively to Bernoulli trials of probability of success $p \in]0, 1[$. The number of repetitions of a Bernoulli experiment of parameter

p which is necessary to obtain k successes is said to follow a Negative Binomial random variable with parameters k and p , denoted by $X_k \sim \mathcal{NB}(k, p)$. For $k = 1$, it is said that X_1 follows a geometric law with parameter p , denoted $X_1 \sim \mathcal{G}(p)$.

Similarly to the sequence of binomial random variable, we may apply the central limit theorem to the sequence of negative binomial random variables X_k , $k \geq 1$ to get the following result

$$(3.3) \quad Z_n = \frac{p(X_k - \frac{k}{p})}{\sqrt{nq}} \rightsquigarrow \mathcal{N}(0, 1) \text{ as } k \rightarrow +\infty.$$

To this purpose, the reader may find more details in classical elementary books in probability theory, for instance in [6] or in [7], Chapters 2 and 3. In Chapter 2 of these monographs, Lemma 2, ensures that a $\mathcal{NB}(k, p)$ random variable X_k is the sum of k independent and geometric $\mathcal{G}(p)$ random variables Y_1, \dots, Y_k such that

$$X_k = Y_1 + \dots + Y_n,$$

and for each of these random variables Z_i 's, we have

$$\mathbb{E}(Y_i) = \frac{1}{p} \text{ and } \sigma^2 = \mathbb{V}ar(Y_i) = \frac{q}{p^2} \text{ where } q = 1 - p.$$

Thus, by the central limit theorem

$$Z_n = \frac{p(X_k - \frac{k}{p})}{\sqrt{nq}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)) \rightsquigarrow \mathcal{N}(0, 1) \text{ as } k \rightarrow +\infty.$$

which proves (3.3).

3.6. Limit laws in Extreme value Theory. Consider X_1, X_2, \dots a sequence of *iid* random variables with common distribution function F . Put for each $n \geq 1$,

$$M_n = \max(X_1, \dots, X_n).$$

Recall that for any $x \in \mathbb{R}$

$$P(M_n \leq x) = F(x)^n, x \in \mathbb{R}.$$

The basic problem of extreme value theory is finding sequences $(a_n > 0)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ such that

$$\frac{M_n - b_n}{a_n}$$

weakly converges to some random variable Z ,

$$\frac{M_n - b_n}{a_n} \rightsquigarrow Z.$$

If this holds, we write $F \in D(F_Z)$.

We are going to give three examples corresponding to the three non-trivial cases.

(a) Let Λ be a Gumbel random variable of distribution function

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}.$$

Let the X_i 's are standard exponential random variables, $X_i \sim \mathcal{E}(1)$, with

$$F(x) = (1 - \exp(-x))1_{(x \geq 0)}, \quad x \in \mathbb{R}.$$

We have, as $n \rightarrow +\infty$,

$$(3.4) \quad M_n - \log n \rightsquigarrow \Lambda.$$

Proof. By using the distribution functions, we want to prove that for any $x \in \mathbb{R}$,

$$(3.5) \quad \mathbb{P}(M_n - \log n \leq x) \rightarrow \Lambda(x).$$

Indeed, we have

$$\mathbb{P}(M_n - \log n \leq x) = P(M_n \leq x + \log n) = F(x + \log n)^n.$$

But for any $x \in \mathbb{R}$, $x + \log n \geq 0$ for $n \geq \exp(-x)$. Then for large values of n , $P(M_n \leq x + \log n) = (1 - \exp(-x - \log n))$ and next for any $x \in \mathbb{R}$ and for n large enough,

$$\mathbb{P}(M_n - \log n \leq x) = \left(1 - \frac{e^{-x}}{n}\right) \rightarrow e^{-e^{-x}} = \Lambda(x).$$

So (3.5) holds and so does (3.4), that is : $X \in D(\Lambda)$.

(b) Let $FR(\alpha)$ a Fréchet random variable with parameter $\alpha > 0$, with distribution function

$$\phi_\alpha(x) = \exp(-x^{-\alpha})1_{(x \geq 0)},$$

where 1_A is the indicator function of the set A that assigns the value one to elements of A and zero to elements of the complementary of A .

Let the X_i 's be Pareto random variables with parameter $\alpha > 0$, $X \sim \mathcal{Par}(\alpha)$, with common distribution functions

$$F(x) = (1 - x^{-\alpha}) 1_{(x \geq 1)}, \quad x \in \mathbb{R}$$

Then, as $n \rightarrow +\infty$, we have

$$(3.6) \quad n^{-1/\alpha} M_n \rightsquigarrow FR(\alpha).()$$

Proof. We want to prove that for any $x \in \mathbb{R}$, we have as $n \rightarrow +\infty$,

$$(3.7) \quad \mathbb{P}(n^{-1/\alpha} M_n \leq x) \rightarrow \phi_\alpha(x).$$

The observations X_i 's are nonnegative since the support of a $\mathcal{Par}(\alpha)$ law is \mathbb{R}_+ . So the maxima M_n are nonnegative for any $n \geq 1$. We may discuss two cases.

Case $x \leq 0$. In this case, we have

$$\mathbb{P}(n^{-1/\alpha} M_n \leq 0) = 0 = \phi_\alpha(x),$$

and then (3.7) holds.

Case $x > 0$. In this case

$$P(n^{-1/\alpha} M_n \leq x) = P(M_n \leq n^{1/\alpha} x).$$

For large values of n , we have $n^{1/\alpha} x > 1$ (take $n \geq x^{-\alpha}$, to ensure that) and for these values,

$$\begin{aligned} \mathbb{P}(n^{-1/\alpha} M_n \leq x) &= F(n^{1/\alpha} x)^n = (1 - (n^{1/\alpha} x)^{-\alpha})^n \\ &= \left(1 - \frac{x^{-\alpha}}{n}\right)^n \rightarrow \exp(-x^{-\alpha}) = \phi_\alpha(x). \end{aligned}$$

So (3.7) holds for $x > 0$. But putting together the two cases, we have $F \in D(FR(\alpha))$.

(c) Let $W(\beta)$ be a Weibull random variable with parameter $\beta > 0$, with distribution function

$$\psi_\alpha(x) = \exp(-(-x)^\beta)1_{x \leq 0} + 1_{(x > 0)}.$$

Let the X_i 's be uniformly distributed on $(0, 1)$ with probability distribution function :

$$F(x) = x1_{(0 \leq x \leq 1)} + 1_{(x \geq 1)}, \quad x \in \mathbb{R}.$$

We have

$$(3.8) \quad n(M_n - 1) \xrightarrow{d} W(1) \text{ as } n \rightarrow +\infty.$$

Proof. We have to prove that for any $x \in \mathbb{R}$, as $n \rightarrow +\infty$,

$$(3.9) \quad \mathbb{P}(n(M_n - 1) \leq x) = F\left(1 + \frac{x}{n}\right)^n \rightarrow \psi_1(x).$$

We have two cases.

Case $x \geq 0$. We see that $1 + x/n$ is nonnegative for $n \geq 1$ and

$$P(n(M_n - 1) \leq x) = F\left(1 + \frac{x}{n}\right)^n = 1 = \psi_1(x)$$

and we see that (3.9) holds for $x \geq 0$.

Case $x < 0$. For large values of n , we have $0 \leq 1 + x/n \leq 1$ (take $0 \geq -x \geq n$, to get it) and for these values of n ,

$$\begin{aligned} P(n(M_n - 1) \leq x) &= F\left(1 + \frac{x}{n}\right)^n \\ &= \left(1 + \frac{x}{n}\right)^n \rightarrow e^x = \psi_1(x). \end{aligned}$$

Then (3.9) also holds for $x < 0$ and then (3.9) holds for any $x \in \mathbb{R}$,

$$\mathbb{P}(n(M_n - 1) \leq x) \longrightarrow \psi_1(x).$$

Conclusion : $F \in D(W(1))$.

Summary. In Univariate Extreme Value Theory (UEVT), it is proved that the three nondegenerated possible limits are the three we gave above. You will have the opportunity to go deep in that theory in the book of this series [5].

4. Examples of Convergence in \mathbb{R}^k

4.1. Simple Central Limit in \mathbb{R}^k . We now move to the Central Limit Theorem in \mathbb{R}^k in the *iid* case. Let X_1, X_2, \dots be centered *iid* \mathbb{R}^k -random variables with common finite variance-covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i \leq k, 1 \leq j \leq k}$, that is

$$\sigma_{ij} = \text{Cov}(X_i, X_j) \in \mathbb{R}, \quad 1 \leq i, j \leq k.$$

Set the partial sums

$$S_n = X_1 + X_2 + \dots + X_n, \quad n \geq 1.$$

We have the following central limit theorem on \mathbb{R}^k ,

$$S_n/\sqrt{n} \rightsquigarrow \mathcal{N}(0, \Sigma) \quad \text{as } n \rightarrow +\infty.$$

Proof. The matrix Σ is symmetrical and nonnegative in the sense that for any $u \in \mathbb{R}^k$

$${}^t u \Sigma u = {}^t u \mathbb{E}(X X') u = \mathbb{E}(({}^t X u)({}^t X u)) = \mathbb{E}(({}^t X u)^2) \geq 0.$$

By the matrices theory, Σ has k nonnegative eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ and there exists a orthogonal (k, k) -matrix T such that

$${}^t T \Sigma T = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \Lambda.$$

Set

$$Y_i = {}^t T X_i.$$

The random variables Y_i are centered, *iid* and have common variance-covariance matrix equal to

$$\Sigma_Y = {}^t T \Sigma T = \Lambda.$$

This means that the components of each Y_i are uncorrelated and have variances equal to $\lambda_1, \lambda_2, \dots, \lambda_n$. Set

$$(4.1) \quad M_n = \frac{1}{\sqrt{n}}(Y_1 + Y_2 + \dots + Y_n) = {}^t T \left(\frac{S_n}{\sqrt{n}} \right).$$

For any $A = {}^t(a_1, a_2, \dots, a_k) \in \mathbb{R}^k$,

$$\langle A, M_n \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n \langle A, Y_i \rangle.$$

The variables $\langle A, Y_i \rangle$ then are *iid* and have common variance

$$\mathbb{E} \langle A, Y_i \rangle^2 = \sum_{i=1}^n a_i^2 \lambda_i = {}^t A \Lambda A,$$

because of the uncorrelation of the components of each Y_i . We may apply the central limit theorem in \mathbb{R} to get

$$\langle A, M_n \rangle \rightarrow \mathcal{N}(0, \sum_{i=1}^n a_i^2 \lambda_i) = \mathcal{N}(0, {}^t A \Lambda A)$$

But $\mathcal{N}(0, {}^t A \Lambda A)$ is the law of a Gaussian random variable that is the linear transform ${}^t A Z = \langle A, Z \rangle$ of Z , where Z follows the $\mathcal{N}(0, \Lambda)$ law. Then

$$\forall A \in \mathbb{R}^k, \langle A, M_n \rangle \rightsquigarrow \langle A, Z \rangle.$$

In terms of characteristic functions, we have for any $t \in \mathbb{R}$ and for any $A \in \mathbb{R}^k$,

$$\mathbb{E} \exp(it \langle A, M_n \rangle) \rightarrow \mathbb{E} \exp(it \langle A, Z \rangle).$$

For $t = 1$, we have for any $A \in \mathbb{R}^k$

$$\Phi_{M_n}(A) = \mathbb{E} \exp(i \langle A, M_n \rangle) \rightarrow \Phi_Z(A) = \mathbb{E} \exp(i \langle A, Z \rangle).$$

This means that

$$M_n \rightsquigarrow Z.$$

This, Point (e) of Theorem 1 and (4.1) together implies that

$$S_n / \sqrt{n} = T M_n \rightsquigarrow T Z$$

and then

$${}^t T Z \sim \mathcal{N}(0, T \Lambda {}^t T) = \mathcal{N}(0, \Sigma).$$

Hence, finally, as $n \rightarrow +\infty$,

$$S_n / \sqrt{n} \rightsquigarrow \mathcal{N}(0, \Sigma)$$

4.2. Weak Convergence of the Multinomial Law. A k -tuple $X_n = (X_{1,n}, \dots, X_{k,n})$ follows a multinomial law with parameters $n \geq 1$ and $p = (p_1, p_2, \dots, p_k)$, with

$$\forall (1 \leq i \leq k), p_i > 0 \text{ and } \sum_{1 \leq i \leq k} p_i = 1,$$

denoted $X \sim \mathcal{M}_k(n, p)$, if and only if its probability law is given by

$$\mathbb{P}(X_{1,n} = n_1, \dots, X_{k,n} = n_k) = \frac{n!}{n_1! \times \dots \times n_k!} p_1^{n_1} \times p_2^{n_2} \times \dots \times p_k^{n_k},$$

where (n_1, \dots, n_k) satisfies

$$\forall (1 \leq i \leq k), n_i \geq 0 \text{ et } \sum_{1 \leq i \leq k} n_i = n.$$

A random variable following the $\mathcal{M}_k(n, p)$ law may be generated as follows :

Consider a random experiment with k possible outcomes $E_i, 1 \leq i \leq k$, each of them occurring with $p_i > 0$. After n repetitions, the number of occurrences $X_{i,n}$ of each E_i is observed for $i = 1, \dots, k$. The resulting random vector follows the $\mathcal{M}_k(n, p)$ law. Each individual coordinate $X_{i,n}$ follows the Binomial law $\mathcal{B}(n, p_i)$.

We have the following weak convergence result.

Put

$$(4.2) \quad Z_n = {}^t \left(\frac{X_{1,n} - np_1}{\sqrt{np_1}}, \dots, \frac{X_{k,n} - np_k}{\sqrt{np_k}} \right) \rightsquigarrow \mathcal{N}_k(0, \Sigma) \text{ as } n \rightarrow +\infty,$$

where Σ is a (k, k) -matrix with elements $\Sigma_{i,i} = 1 - p_i$ and $\Sigma_{i,j} = \sqrt{p_i p_j}$, $1 \leq i, j \leq k$.

Important remark. This result has a significant number of applications. We may cite two of them. It is used to have the finite-distribution function of the empirical process. It also serves as the foundations of Chi-square statistical tests that will be studied latter in one the books of this series.

Proof. We have at least two ways of proving the result. The first is based on the use of the moment function tool and on logarithm expansions. The second exploits the central limit theorem in \mathbb{R}^k we just

proved.

First proof. We already know from [9] that its moment generating function is

$$\phi_{X_n}(u) = \left(\sum_{1 \leq i \leq k} p_i e^{u_i} \right)^n.$$

We have

$$Z_n = AX + B,$$

where A is the diagonal matrix

$$A = \text{diag} \left(\frac{1}{\sqrt{np_1}}, \frac{1}{\sqrt{np_2}}, \dots, \frac{1}{\sqrt{np_k}} \right)$$

and

$$B = \begin{pmatrix} -\sqrt{np_1} \\ -\sqrt{np_2} \\ \dots \\ -\sqrt{np_k} \end{pmatrix}.$$

Thus

$$\begin{aligned} \phi_{Z_n}(u) &= \exp(\langle B, u \rangle) \times \phi_X({}^t A u) \\ &= \left(\exp \left(\sum_{1 \leq i \leq k} -\sqrt{np_i} u_i \right) \right) \times \left(\sum_{1 \leq i \leq k} p_i e^{u_i / \sqrt{np_i}} \right)^n. \end{aligned}$$

Let u be fixed. For each fixed i , $1 \leq i \leq k$, $u_i / \sqrt{np_i} \rightarrow +\infty$ as $n \rightarrow \infty$ since $p_i > 0$. We have the expansion

$$e^{u_i / \sqrt{np_i}} = 1 + u_i / \sqrt{np_i} + \frac{1}{2} \frac{u_i^2}{np_i} + O(n^{-3/2}).$$

Next

$$\begin{aligned} A &= \left(\sum_{1 \leq i \leq k} p_i e^{u_i / \sqrt{np_i}} \right)^n = \exp \left(n \log \left(\sum_{1 \leq i \leq k} p_i e^{u_i / \sqrt{np_i}} \right) \right) \\ &= \exp \left(n \log \left(1 + \sum_{1 \leq i \leq k} u_i \sqrt{p_i / n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} + O(n^{-3/2}) \right) \right). \end{aligned}$$

Set

$$a = \sum_{1 \leq i \leq k} u_i \sqrt{p_i / n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We have

$$A = \exp(n \log(1 + a)).$$

Let us expand $\log(1 + a)$ at the second order 2. We obtain

$$\begin{aligned}
A &= \exp\left(n\left(a - \frac{1}{2}a^2 + O(a^3)\right)\right) \\
&= \exp\left(n\left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i/n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i/n}\right)^2 + O(n^{-3/2})\right)\right) \\
&= \exp\left(\sum_{1 \leq i \leq k} u_i \sqrt{np_i} + \sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right) \\
&= \exp\left(\sum_{1 \leq i \leq k} u_i \sqrt{np_i}\right) \times \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right).
\end{aligned}$$

Putting all this together, we get

$$\begin{aligned}
\phi_{Z_n}(u) &= \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right) \\
&\rightarrow \phi_Z(u) = \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2\right).
\end{aligned}$$

and

$$(4.3) \quad \phi_Z(u) = \exp\left\{\sum_{1 \leq i \leq k} \frac{1}{2} (1 - p_i) u_i^2 - \sum_{1 \leq i, j \leq k} u_i u_j \sqrt{p_i p_j}\right\},$$

which is the moment function of a k -dimensional centered Gaussian vector Z whose variance-covariance matrix is Σ . The first proof finishes here.

Second proof. At the i -th repetition of the experiment, $i \in \{1, \dots, n\}$, we have a random vecteur

$$Z^{(i)} = \begin{pmatrix} Z_1^{(i)} \\ \dots \\ Z_k^{(i)} \end{pmatrix}$$

defined as follows : for each $1 \leq r \leq k$

$$Z_r^{(i)} = \begin{cases} 1 & \text{if the outcome } E_r \text{ occurs at the } i^{th} \text{ experiment and any other did not} \\ 0 & \text{if a different outcome occurs at the } i^{th} \text{ experiment} \end{cases}$$

It is clear that each $Z^{(i)}$ is distributed as a multivariate $\mathcal{M}_k(1, k)$ random variable, and that the $Z^{(i)}$'s are independent.

Further, for a fixed $i \in \{1, \dots, n\}$, each $Z_r^{(i)}$, $1 \leq r \leq k$, follows a Bernoulli law of parameter p and only one of the $Z_r^{(i)}$'s ($1 \leq r \leq k$) takes the value one, the others being null. This implying that

$$Z_r^{(i)} Z_s^{(i)} = 0 \text{ for } 1 \leq r \neq s \leq k, 1 \leq i \leq n.$$

We also have

$$Z_1^{(i)} + \dots + Z_k^{(i)} = 1.$$

Then for each $i \in \{1, \dots, n\}$,

$$\mathbb{E}(Z_r^{(i)}) = p_i \text{ and } \mathbb{V}ar(Z_r^{(i)}) = p_i(1 - p_i), \quad 1 \leq r \leq k$$

and for $1 \leq r \neq s \leq k$

$$\text{cov}(Z_r^{(i)}, Z_s^{(i)}) = \mathbb{E}(Z_r^{(i)} Z_s^{(i)}) - \mathbb{E}(Z_r^{(i)})\mathbb{E}(Z_s^{(i)}) = -p_r p_s,$$

since $Z_r^{(i)} Z_s^{(i)} = 0$. So, each $Z^{(i)}$ has the variance-covariance matrix

$$\Sigma^0 = \begin{pmatrix} p_1(1 - p_1) & -p_1 p_2 & \dots & -p_1 p_{k-1} & -p_1 p_k \\ -p_2 p_1 & p_2(1 - p_2) & \dots & -p_2 p_{k-1} & -p_2 p_k \\ \dots & \dots & \dots & \dots & \dots \\ -p_{k-1} p_1 & -p_{k-1} p_2 & \dots & p_{k-1}(1 - p_{k-1}) & -p_{k-1} p_k \\ -p_k p_1 & -p_k p_2 & \dots & -p_k p_{k-1} & -p_k(1 - p_k) \end{pmatrix}$$

or, in a different notation,

$$\Sigma_0 = (\sigma_{ij}^0)_{1 \leq i, j \leq k} \text{ with } \sigma_{ij}^0 = \begin{cases} p_i(1 - p_i) & \text{if } i = j \\ -p_i p_j & \text{if } i \neq j \end{cases}.$$

After n repetitions of the experiment, the random variables $Z^{(1)}, \dots, Z^{(n)}$ which are independent and $\mathcal{M}_k(1, k)$ random vectors add up to X_n , which means that

$$X_n = Z^{(1)} + \dots + Z^{(n)}.$$

By the multivariate standard central limit theorem, we have

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z^{(i)} - \mathbb{E}(Z^{(i)})) \rightsquigarrow Z_0 \sim \mathcal{N}_k(0, \Sigma_0).$$

We easily check that

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z^{(i)} - \mathbb{E}(Z^{(i)})) = {}^t \left(\frac{X_{1,n} - np_1}{\sqrt{n}}, \frac{X_{2,n} - np_2}{\sqrt{n}}, \dots, \frac{X_{k,n} - np_k}{\sqrt{n}} \right).$$

And then, we have the matrix relation

$$DS_n = Z_n,$$

where D is the diagonal matrix

$$D = \text{diag}(1/\sqrt{p_1}, \dots, 1/\sqrt{p_k}).$$

By the continuous mapping theorem (Point (e) of Theorem 1),

$$Z_n = DS_n \rightsquigarrow DZ_0 \sim \mathcal{N}_k(0, D\Sigma_0 D),$$

since D is a symmetrical matrix. It remains to compute

$$\Sigma = D\Sigma_0 D = (\sigma_{ij})_{1 \leq i, j \leq k}.$$

For $1 \leq h, j \leq k$, $(\Sigma_0 D)_{hj}$ is the matrix product of the h^{th} row of Σ_0 by the j^{th} column of D . By using the fact that D is diagonal, we get for $1 \leq h, j \leq k$,

$$(\Sigma_0 D)_{hj} = \sigma_{hj}^0 / \sqrt{p_j}.$$

Next $\sigma_{ij} = (D\Sigma_0 D)_{ij}$ is the product of i^{th} row of D by the j^{th} column of $(\Sigma_0 D)^{(j)} = {}^t((\Sigma_0 D)_{1j}, (\Sigma_0 D)_{2j}, \dots, (\Sigma_0 D)_{kj})$ and then, by using the diagonal property of D , we have

$$(D\Sigma_0 D)_{ij} = \frac{1}{\sqrt{p_i}} (\Sigma_0 D)_{ij},$$

and then

$$\sigma_{ij} = (D\Sigma_0 D)_{ij} = \frac{1}{\sqrt{p_i p_j}} \sigma_{ij}^0 = \begin{cases} \sigma_{ii}^0 / p_i = 1 - p_i & \text{if } i = j \\ -\sqrt{p_i p_j} & \text{if } i \neq j \end{cases} \quad \dots$$

We get again that

$$Z_n \rightsquigarrow \mathcal{N}_k(0, \Sigma),$$

where Σ is defined in the line following Formula (4.2) in head part of this subsection. This ends the second proof.

We may conclude in a form of a proposition.

PROPOSITION 2. *Let $X(n) = (X_1(n), \dots, X_k(n))$, $n \geq 1$, be a sequence of k -dimensional random vectors such that each $X(n)$ follows a multinomial law with parameters $n \geq 1$ and $p = (p_1, p_2, \dots, p_k)$ with*

$$\forall (1 \leq i \leq k), p_i > 0 \quad \text{and} \quad \sum_{1 \leq i \leq k} p_i = 1.$$

Then, as $n \rightarrow +\infty$,

$$Z_n = {}^t\left(\frac{X_1 - np_1}{\sqrt{np_1}}, \dots, \frac{X_1 - np_k}{\sqrt{np_k}}\right)$$

weakly converges to a k -dimensional Gaussian vector of variance-covariance matrix Σ whose elements are

$$(4.4) \quad \Sigma_{ii} = (1 - p_i)$$

and

$$(4.5) \quad \Sigma_{ij} = -\sqrt{p_i p_j}.$$

4.3. Finite dimensional weak limits of the uniform empirical process. Let U_1, U_2, \dots be a sequence of independent and standard uniformly distributed random variables on $(0, 1)$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with common distribution function $F(s) = s1_{(0 \leq s \leq 1)} + 1_{(s \geq 1)}$. For each $n \geq 1$, we may define the empirical distribution function associated with U_1, U_2, \dots, U_n :

$$\mathbb{R} \ni x \mapsto U_n(s) = \frac{1}{n} \text{Card}\{i, 1 \leq i \leq n, U_i \leq s\}$$

The empirical process associated with U_1, U_2, \dots, U_n is defined as follows

$$\alpha_n(s) = \sqrt{n}(U_n(s) - s), 0 \leq s \leq 1.$$

Consider $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = 1$ a partition of $(0, 1)$ and set

$$Y_n = {}^t(\alpha_n(t_1), \dots, \alpha_n(t_{k+1})).$$

We have :

PROPOSITION 3. *Any finite distribution of the uniform empirical process of the form*

$${}^t(\alpha_n(t_1), \dots, \alpha_n(t_{k+1}))$$

with

$$0 = t_0 < t_1 < \dots < t_k < t_{k+1} = 1,$$

weakly converges to a k -dimensional Gaussian random variable with variance-covariance matrix

$$(\min(t_i, t_j) - t_i t_j)_{1 \leq i, j \leq k},$$

that is,

$$(\alpha_n(t_1), \dots, \alpha_n(t_{k+1})) \rightarrow \mathcal{N}_k(0, (\min(t_i, t_j) - t_i t_j)_{1 \leq i, j \leq k})$$

Proof. Set

$$(4.6) \quad Z_n = {}^t \left(\frac{\alpha_n(t_1)}{\sqrt{t_1}}, \frac{\alpha_n(t_2) - \alpha_n(t_1)}{\sqrt{t_2 - t_1}}, \dots, \frac{\alpha_n(t_{k+1}) - \alpha_n(t_k)}{\sqrt{t_{k+1} - t_k}} \right).$$

Let us remark that

$$N_n = (nF_n(t_1), nF_n(t_2) - nF_n(t_1), \dots, nF_n(t_{k+1}) - nF_n(t_k))$$

follows a multinomial law with outcomes probabilities $t_1, t_2 - t_1, \dots, t_{k+1} - t_k$. Indeed, we have that

$$nF_n(t_j) - nF_n(t_{j-1})$$

is the number of observations falling in $]t_{j-1}, t_j]$ and for each j , the probability that one observation falls in $]t_{j-1}, t_j]$ is $p_j = t_j - t_{j-1}$.

We may apply the weak convergence of the multinomial law we established in Subsection 4.2.

Let us define Z_n by centering each j th component of N_n at $n(t_j - t_{j-1})$ and normalizing it by $\sqrt{n(t_j - t_{j-1})}$.

Remind that $Y_n = {}^t(\alpha_n(t_1), \dots, \alpha_n(t_{k+1}))$. We have the matrix relation

$$Z_n = AY_n \Leftrightarrow Y_n = BZ_n$$

where the relation $y = Bz$ is the following correspondance

$$y_i = \sqrt{t_1}x_1 + \sqrt{t_2 - t_1}x_2 + \dots + \sqrt{t_2 - t_1}x_i, \quad 1 \leq i \leq k+1.$$

By the weak convergence of the multinomial law, Z_n weakly converges to a centered Gaussian vector $Z = (Z_1, Z_2, \dots, Z_{k+1})$ such that

$$\mathbb{E}(Z_j^2) = 1 - (t_j - t_{j-1})$$

and

$$\mathbb{E}(Z_i Z_j) = -\sqrt{(t_i - t_{i-1})(t_j - t_{j-1})}.$$

By the continuous mapping theorem (Point (e) of Theorem 1), $Y_n = BZ_n$ weakly converges to $Y = BZ$, where

$$Y_i = \sqrt{t_1}Z_1 + \sqrt{t_2 - t_1}Z_2 + \dots + \sqrt{t_2 - t_1}Z_i, \quad 1 \leq i \leq k+1.$$

Let $T = (T_1, \dots, T_{k+1})$ be defined by $(t_j - t_{j-1})Z_j = T_j$, $1 \leq j \leq k$, that is

$$Z = {}^t \left(\frac{T_1}{\sqrt{t_1}}, \frac{T_2}{\sqrt{(t_2 - t_1)}}, \dots, \frac{T_j}{\sqrt{(t_j - t_{j-1})}}, \dots, \frac{T_{k+1}}{\sqrt{(t_{k+1} - t_k)}} \right)$$

We have

$$\mathbb{E}(T_j^2) = \mathbb{E} \left(\left(Z_j \sqrt{(t_j - t_{j-1})} \right)^2 = (t_j - t_{j-1})(1 - (t_j - t_{j-1})) \right).$$

and

$$\mathbb{E}(T_i T_j) = \sqrt{(t_j - t_{j-1})(t_i - t_{i-1})} \mathbb{E}(Z_i Z_j) = -(t_j - t_{j-1})(t_i - t_{i-1}).$$

Before we compute the covariance of Y_i and Y_j , we check that for $t_i \leq t_j$,

$$\begin{aligned} t_i t_j &= \left(\sum_{h=1}^i (t_h - t_{h-1}) \right) \left(\sum_{r=1}^j (t_r - t_{r-1}) \right) \\ &= \left(\sum_{h=1}^i (t_h - t_{h-1}) \right) \left(\sum_{r=1}^i (t_r - t_{r-1}) + \sum_{r=i+1}^j (t_r - t_{r-1}) \right) \\ &= \left(\sum_{h=1}^i (t_h - t_{h-1}) \right)^2 + \sum_{h=1}^i \sum_{r=i+1}^j (t_h - t_{h-1})(t_r - t_{r-1}) \\ &= \sum_{h=1}^i (t_h - t_{h-1})^2 + \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\ &\quad - \sum_{h=1}^{h=i} \sum_{r=i+1}^r (t_h - t_{h-1})(t_r - t_{r-1}) \end{aligned}$$

By putting together all these points, we are going to compute the variance-covariance matrix of Y . For $1 \leq i \leq j \leq 1$, we have

$$\begin{aligned} Y_i Y_j &= \left(\sum_{h=1}^i T_h \right)^2 + \sum_{h=1}^{h=i} \sum_{r=i+1}^j T_h T_r \\ &= \sum_{h=1}^i T_h^2 + \sum_{1 \leq h \neq r \leq i} T_h T_r + \sum_{h=1}^i \sum_{r=i+1}^j T_h T_r. \end{aligned}$$

Finally, we get

$$\begin{aligned}
\mathbb{E}(Y_i Y_j) &= \sum_{h=1}^i (1 - (t_h - t_{h-1})) - \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&\quad - \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= \sum_{h=1}^i (t_h - t_{h-1}) - \sum_{h=1}^i (t_h - t_{h-1})^2 \\
&\quad - \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) - \sum_{h=1}^i \sum_{r=i+1}^j (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= t_i - \sum_{h=1}^i (t_h - t_{h-1})^2 - \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&\quad - \sum_{h=1}^i \sum_{r=i+1}^j (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= t_i - t_i t_j = \min(t_i, t_j) - t_i t_j.
\end{aligned}$$

This completes the proof.

5. Invariance principle

Let X_1, X_2, \dots be a sequence of *iid* centered random variables with finite variances, that is $E|X_i|^2 < \infty$. For each $n \geq 1$, set

$$S_n = X_1 + \dots + X_n$$

For $0 \leq t \leq 1$ and $n \geq 1$, put

$$S_n(t) = \frac{S_{[nt]}}{\sqrt{n}}$$

where, for any real u , $[u]$ stands for the integer part of u , which is the greatest integer less or equal to u .

we are going to explore the weak convergence of the finite distributions if the stochastic process $\{S_n(t), 0 \leq t \leq 1\}$.

For this purpose, let $0 = t_0 < t_2 < \dots < t_k = 1$, $k \geq 1$. We have :

PROPOSITION 4. *The sequence of finite distributions*

$$\left(\frac{S_{[nt_j]}}{\sqrt{n}}, 1 \leq j \leq k \right), \quad n \geq 1,$$

weakly converges to k -dimensional centered Gaussian vector with variance-covariance matrix

$$(\min(t_i, t_j))_{1 \leq i, j \leq k}.$$

Proof. We have

$$\begin{cases} Y_n(t_1) = X_n(t_1) - X_n(t_0) = \frac{1}{\sqrt{n}} \sum_{[nt_0] < j \leq [nt_1]} X_j \\ \vdots \\ Y_n(t_i) = X_n(t_i) - X_n(t_{i-1}) = \frac{1}{\sqrt{n}} \sum_{[nt_{i-1}] < j \leq [nt_i]} X_j \\ \vdots \\ Y_n(t_k) = X_n(t_k) - X_n(t_{k-1}) = \frac{1}{\sqrt{n}} \sum_{[nt_{k-1}] < j \leq [nt_k]} X_j \end{cases}.$$

We easily see that the random variables $Y_n(t_i)$ are independent and that for each $1 \leq i \leq k$, we apply the central limit theorem in \mathbb{R} to get,

$$Y_n(t_i) = \frac{1}{\sqrt{n}} \sum_{[nt_{i-1}] < j \leq [nt_i]} X_j \rightarrow \mathcal{N}(0, t_i - t_{i-1})$$

Hence, for any $u = (u_1, \dots, u_k) \in \mathbb{R}^k$,

$$\mathbb{E} \left(\exp \left(\sum_{1 \leq i \leq k} Y_n(t_i) u_i \right) \right) = \prod_{1 \leq i \leq k} \mathbb{E}(\exp(Y_n(t_i) u_i)) \rightarrow \prod_{1 \leq i \leq k} e^{\frac{1}{2} u_i^2 / (t_i - t_{i-1})}.$$

Thus, the vector $Y_n = (Y_n(t_i), 1 \leq i \leq k)$ weakly converges to a Gaussian random vector Z , which has independent components and for each $1 \leq i \leq k$, the i -th component has the variance $t_i - t_{i-1}$.

The vector $X_n = (X_n(t_i), 1 \leq i \leq k)$ is the linear transform of Y_n of the form

$$X_n = AY_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ 1 & \dots & 1 & 0 \\ 1 & \dots & \dots & 1 \end{pmatrix} Y_n$$

with

$$A_{ij} = 1_{(i \leq j)}.$$

Then X_n weakly converges to $V = AZ$, whose components satisfy

$$V_i = Z_1 + \dots + Z_i$$

and

$$Z_i = V_i - V_{i-1}.$$

Then for any $1 \leq i \leq k$,

$$\mathbb{E}(V_i^2) = \sum_{1 \leq j \leq i} \mathbb{E}(Z_j^2) = \sum_{1 \leq j \leq i} (t_j - t_{j-1}) = t_i.$$

And for any $1 \leq i \leq j \leq k$,

$$\begin{aligned} \mathbb{E}(V_i V_j) &= \mathbb{E}(V_i(V_i + (V_j - V_i))) \\ &= \mathbb{E}(V_i^2) + \mathbb{E}(V_i(V_j - V_i)). \end{aligned}$$

Since

$$V_i = Z_1 + \dots + Z_i$$

and since the random variables

$$Z_i = V_i - V_{i-1}$$

are independent and centered, we get

$$\mathbb{E}(V_i V_j) = \mathbb{E}(V_i^2) = t_i = t_i \wedge t_j.$$

This suffices to conclude.

Terminology. The result presented in Proposition 4 is the first step of what is called *invariance principle* in Probability Theory.

CHAPTER 2

Weak Convergence Theory

1. Introduction

In this chapter, we treat a unified theory of weak convergence by its functional characterization. We want to have complete theory of limits of sequences of probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, where $\mathcal{B}(\mathbb{R}^k)$ is the Borel σ -algebra of \mathbb{R}^k .

However, the handling of the fundamental results only uses the metric structure of \mathbb{R}^k . This is why, whenever possible, we deal with sequences of probability measures on a metric spaces (S, d) , endowed with its Borel σ -algebra $\mathcal{B}(S)$.

But when dealing with limits of subsequences of sequences of random variables or probability measures, we essentially remain in \mathbb{R}^k by making profit of the Helly-Bray theorem.

As in any limit theory, we will have to deal with the unicity of limits, and convergence criteria, and relative compactity. Here, we will speak of weak compactity or simply tightness or uniform tightness.

2. Definition, Unicity and Portmanteau Theorem

DEFINITION 1. *The sequence of measurable applications $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, \mathcal{B}(S))$ weakly converges to the measurable application $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (S, \mathcal{B}(S))$ if and only for any continuous and bounded function $f : S \mapsto \mathbb{R}$, (denoted $f \in \mathcal{C}_b(S)$), we have*

$$(2.1) \quad \mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X) \text{ as } n \rightarrow +\infty.$$

We notice that the spaces on which the applications X_n are defined have no importance here. Only matter their probability laws on (S, d) . Indeed, denote $L = \mathbb{P}_X = \mathbb{P}_\infty \circ X^{-1}$, the probability law of X defined by

$$\forall B \in \mathcal{B}(S), \quad L(B) = \mathbb{P}_\infty(X^{-1}(B)) = \mathbb{P}_\infty(X \in B).$$

and for each $n \geq 1$, $\mathbb{P}^{(n)}$ the probability law of X_n defined by

$$\forall B \in \mathcal{B}(S), \mathbb{P}^{(n)}(B) = \mathbb{P}_n(X_n^{-1}(B)) = \mathbb{P}_n(X_n \in B).$$

The definition says that X_n weakly converges to X if and only if for any $f \in \mathcal{C}_b(S)$,

$$\int_S f(x) d\mathbb{P}^{(n)}(x) \rightarrow \int_S f(x) dL(x) \text{ as } n \rightarrow +\infty.$$

We might also replace (2.1) by

$$(2.2) \quad \mathbb{E}f(X_n) \rightarrow \int_S f dL \text{ as } n \rightarrow +\infty,$$

and only say that $(X_n)_{n \geq 1}$ weakly converges to the probability measure L . In the sequel, we will use both terminologies.

Warning. It is also important to see that the expectation symbols in (2.1) depend of the probability measures that they use, and consequently, they should be labelled accordingly as

$$\mathbb{E}_\infty(f(X)) = \int f(X) d\mathbb{P}_\infty, \quad \mathbb{E}_n(f(X_n)) = \int f(X_n) d\mathbb{P}_n, \quad n \geq 1.$$

But choose to not put the subscripts n and ∞ to keep the writing simple and to use them only when necessary.

Notation. When $(X_n)_{n \geq 1}$ weakly converges X as $n \rightarrow +\infty$, we mainly use the notation

$$X_n \rightsquigarrow X \text{ as } n \rightarrow +\infty,$$

but we may also use $X_n \rightarrow_w X$ (w standing for *weakly*) or $X_n \rightarrow_d X$ (d standing for : *in distribution*).

We are going to show that the limit we have defined is unique, **but in distribution**, in the following sense.

PROPOSITION 5. *Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, \mathcal{B}(S))$ be a sequence of measurable applications and, \mathbb{P}_1 and \mathbb{P}_2 two probability measures on $(S, \mathcal{B}(S))$. Suppose that X_n weakly converges to \mathbb{Q}_1 and to \mathbb{Q}_2 . Then, we necessarily have*

$$\mathbb{Q}_1 = \mathbb{Q}_2.$$

This means that if X_n weakly converges to X and to Y , then they have the same probability measure, meaning that they are equal in distribution.

Proof. Suppose that X_n weakly converges to \mathbb{P}_1 and to \mathbb{P}_2 . We want to show that $\mathbb{Q}_1 = \mathbb{Q}_2$. But it suffices to show that the two probability measures coincide on the class Θ of open sets of (S, d) . Indeed, the class Θ is a π -system (that is : a class which is closed under finite intersection), which generates $\mathcal{B}(S)$. Then, by the $\lambda - \pi$ lemma, two probability measures on $(S, \mathcal{B}(S))$ that coincide on Θ are equal on $\mathcal{B}(S)$.

Now let G be an open set of S . For any integer number $m \geq 1$, set the function $f_m(x) = \min(m d(x, G^c), 1)$, $x \in S$. We may see that for any $m \geq 1$, f_m has values in $[0, 1]$, and is bounded. Since G^c is closed, we have

$$d(x, G^c) = \begin{cases} > 0 & \text{si } x \in G \\ 0 & \text{si } x \in G^c \end{cases}.$$

Let us show that f_m is a Lipschitz function. Let us handle $|f_m(x) - f_m(y)|$ through three cases.

Case 1. $(x, y) \in (G^c)^2$. Then

$$|f_m(x) - f_m(y)| = 0 \leq m d(x, y).$$

Case 2. $x \in G$ and $y \in G^c$ (including also the case where the roles of x and y are switched). We have

$$|f_m(x) - f_m(y)| = |\min(m d(x, G^c), 1)| \leq m d(x, G^c) \leq m d(x, y),$$

by the very definition of $d(x, G^c) = \inf\{d(x, z), z \in G^c\}$.

Case 3. $(x, y) \in G^2$. We use Property (7.5) the Annexe Section (7) below and get,

$$\begin{aligned} |f_m(x) - f_m(y)| &= |\min(m d(x, G^c), 1) - \min(m d(y, G^c), 1)| \leq |m d(x, G^c) - m d(y, G^c)|, \\ &\leq m d(x, y) \end{aligned}$$

by the second triangle inequality. Then f_m is a Lipschitz function with coefficient m . Now, let us show that

$$f_m \uparrow 1_G \text{ as } m \uparrow \infty.$$

Indeed, if $x \in G^c$, we obviously have $f_m(x) = 0 \uparrow 0 = 1_G(x)$. If $x \in G$, that $d(x, G^c) > 0$ and $md(x, G^c) \uparrow \infty$ as $m \uparrow \infty$. Then for m large enough,

$$(2.3) \quad f_m(x) = 1 \uparrow 1_G(x) = 1 \text{ as } m \uparrow \infty.$$

In summary, each function f_m is a nonnegative and bounded Lipschitz function, that implies that $f_m \in \mathcal{C}_b(S)$, $m \geq 1$.

Now let us apply the definition of the weak convergence. The assumption implies that for any $f \in \mathcal{C}_b(S)$, we have as $n \rightarrow +\infty$,

$$(2.4) \quad \mathbb{E}f(X_n) \rightarrow \int f d\mathbb{Q}_1 \text{ and } \mathbb{E}f(X_n) \rightarrow \int f d\mathbb{Q}_2.$$

By unicity of real limits in \mathbb{R} , we get

$$\forall (f \in \mathcal{C}_b(S)), \int f d\mathbb{Q}_1 = \int f d\mathbb{Q}_2.$$

Now, we apply this to the f_m , $m \geq 1$ to say

$$\forall (m \geq 1), \int f_m d\mathbb{Q}_1 = \int f_m d\mathbb{Q}_2.$$

Next, as m increases to $+\infty$, we use (2.3) and apply the Monotone Convergence Theorem to conclude that

$$\int 1_G d\mathbb{Q}_1 = \int 1_G d\mathbb{Q}_2,$$

that is

$$\mathbb{Q}_1(G) = \mathbb{Q}_2(G).$$

Since G is arbitrary fixed, this equality holds for all open sets of S . We conclude that $\mathbb{Q}_1 = \mathbb{Q}_2$.

Next, we need to characterize the weak convergence using several criteria. This will furnish a rich set of tools for establishing weak convergence results.

THEOREM 2. *The sequence of measurable applications $X_n : (\Omega_n, \mathcal{A}_n, P_n) \mapsto (S, \mathcal{B}(S))$ weakly converges to the probability measure L if and only if one of these assertions holds.*

(ii) *For any open set G of S ,*

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq L(G).$$

(iii) For any closed set G of S , we have

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq L(F).$$

(iv) For any inferior semi-continuous and bounded below function f , we have

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \int f dL.$$

(v) For any superior semi-continuous and bounded above function f , we have

$$\limsup_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \int f dL.$$

(vi) For any Borel set B of S that is L -continuous, that is $L(\partial B) = 0$, we have

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in B) = \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = L(B).$$

(vii) For any nonnegative and bounded Lipschitz function f , we have *lipschitzienne*.

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \int f dL.$$

Before we begin the proof, we recall that ∂B is the boundary of the set B . If $L(\partial B) = 0$, it is said that B est L -continuous. As to the semi-continuous function, we will give a reminder in the Annexe below.

Proof. To unify the notation, we denote Formula (2.1) as by Point (i) of the definition of weak convergence. From now, we break the proof into points.

(1) (ii) \Leftrightarrow (iii). This is achieved by complementation.

(2) (iv) \Leftrightarrow (v). This is achieved by moving from f to $-f$ and by remarking that opposite of superior semi-continuous functions are inferior semi-continuous and vice-versa.

(3) (i) \Rightarrow (vii). This is obvious since a Lipschitz function is continuous.

(vii) \Rightarrow (ii). Let G be an open subset of S . For any $m \geq 1$, set $f_m(x) = \min(m d(x, G^c), 1)$. We already knew for the proof of Proposition 5 that for each $m \geq 1$, f_m is a nonnegative and bounded Lipschitz function such that

$$f_m \uparrow 1_G \text{ as } m \uparrow \infty.$$

We have for any $n \geq 1$ and for any $m \geq 1$,

$$\mathbb{E}(1_G(X_n)) \geq \mathbb{E}f_m(X_n).$$

Let us apply (vii) to get

$$(2.5) \quad \liminf_{n \rightarrow +\infty} \mathbb{E}(1_G(X_n)) \geq \liminf_{n \rightarrow +\infty} \mathbb{E}f_m(X_n) \geq \int f_m dL.$$

But for any measurable set B and for any probability measure \mathbb{Q} ,

$$\mathbb{E}_{\mathbb{Q}}(1_B) = \mathbb{Q}(B)$$

For $B = 1_{X_n^{-1}(G)} = 1_{(X_n \in G)}$, we let m increase to $+\infty$ and use the Monotone Convergence Theorem to (2.5), and get

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in G) \geq \int 1_G dL = L(G).$$

Thus (ii) holds true.

(4) (ii) \Rightarrow (iv). Assume (ii) is true. Let f be an inferior semi-continuous function bounded below by M . In a first step, we are going to prove (iv) for $f - M = g$, which is nonnegative and inferior semi-continuous. Then the sets $(g \leq c)$ are closed by Proposition 18 in the Annexe Section 7. Set for $m \geq 1$ fixed,

$$G_i = \{g > i/m\}, \quad i \geq 1$$

and

$$g_m = \frac{1}{m} \sum_{i=1}^{m^2} 1_{G_i}$$

The sets G_i are open since g is inferior semi-continuous. Let us remark that

$$(2.6) \quad g_m(x) = \frac{i}{m} \text{ for } \frac{i}{m} < g(x) \leq \frac{i+1}{m}, \text{ for } i = 1, \dots, m^2 - 1$$

and

$$g_m(x) = m \text{ for } g(x) > m.$$

Then

$$g_m \leq g.$$

Further, by (2.6)

$$|g_m(x) - g(m)| \leq 1/m \text{ pour } g(x) \leq m.$$

This implies

$$g(X_n) \geq g_m(X_n) = \frac{1}{m} \sum_{i=1}^{m^2} 1_{G_i}(X_n) = \frac{1}{m} \sum_{i=1}^{m^2} 1_{(X_n \in G_i)}$$

and next

$$(2.7) \quad \mathbb{E}g(X_n) \geq \mathbb{E}g_m(X_n) = \frac{1}{m} \mathbb{E} \sum_{i=1}^{m^2} 1_{(X_n \in G_i)}.$$

Then (2.7) yields

$$\mathbb{E}g(X_n) \geq \mathbb{E}g_m(X_n) \geq \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{E}1_{(X_n \in G_i)} = \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{P}(X_n \in G_i).$$

By letting n go to $+\infty$ and by applying (ii), we get

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{E}g(X_n) &\geq \liminf_{n \rightarrow +\infty} \mathbb{E}g_m(X_n) \geq \frac{1}{m} \sum_{i=1}^{m^2} L(G_i) = \int g_m \, dL \geq \int_{(g \leq m)} g_m \, dL \\ &\geq \int_{(g \leq m)} g \, dL + \int_{(g \leq m)} (g_m - g) \, dL. \end{aligned}$$

Now, as $m \rightarrow \infty$, we have

$$\int_{(g \leq m)} g \, dL \rightarrow \int g \, dL$$

and next,

$$\left| \int_{(g \leq m)} (g_m - g) \, dL \right| \leq L(S)/m \rightarrow 0.$$

Hence

$$\liminf_{n \rightarrow +\infty} \mathbb{E}g(X_n) \geq \int g \, dL.$$

Now, we come back to f and see that by replacing g by $f - M$ in (vi), the formula remains true for f by simplification of the finite number M . Hence (iv) holds.

(5) (ii) \Rightarrow (vi). Recall that the boundary ∂B of B is the difference of interior B from its adherence (closure), denoted as $\partial B = \overline{B} - \text{int}(B)$. Since

$$\text{int}(B) \subseteq B \subseteq \overline{B},$$

we have

$$(2.8) \quad L(\partial B) = L(\text{int}(B)) - L(\overline{B}) = 0 \Rightarrow L(\text{int}(B)) = L(\overline{B}) = L(B)$$

Since $\text{int}(B)$ is open and \overline{B} is closed, we may apply both (ii) and (iii) to get

$$(2.9) \quad L(\text{int}(B)) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \text{int}(B)) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B),$$

$$(2.10) \quad \leq \mathbb{P}_n(X_n \in \overline{B}) \leq L(\overline{B}),$$

Thus, by (2.8),

$$L(B) = \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B),$$

which was the target.

(6) (vi) \Rightarrow (iii). Assume (vi) holds and let F a closed subset of S . Set $(F_\epsilon) = \{x, d(x, F) \leq \epsilon\}$ for $\epsilon \geq 0$. We have

$$F \subseteq F(\epsilon)$$

and

$$F(\epsilon) \downarrow F \text{ as } \epsilon \downarrow 0$$

Now, $\partial F(\epsilon) \subseteq \{x, d(x, F) = \epsilon\}$ and the sets $\{x, d(x, F) = \epsilon\}$ are disjoint. So the sets $\partial F(\epsilon)$ are disjoint. So they have null probabilities except eventually for a countable number of values of ϵ , that is

$$L\partial F(\epsilon) = 0$$

except eventually for a countable number of values of ϵ . (See Proposition 19 in the Annexe Section 7). Then, we may easily find a sequence $\epsilon_n \downarrow 0$ such that for any $n \geq 1$,

$$L(\partial F(\epsilon_n)) = 0.$$

For n fixed, $F \subseteq F(\epsilon_n)$ and this implies

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F(\epsilon_n))$$

Next, by applying (vi)

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F(\epsilon_n)) = L(F(\epsilon_n))$$

Finally by letting n go to infinity, we arrive at

$$\limsup \mathbb{P}(X_n \in F) \leq L(F),$$

and this is (iii).

(7) (iv) \Rightarrow (i). Assume (iv) is true. Then (v) is also true. Then for any bounded and continuous function f , it is inferior semi-continuous and bounded below and superior semi-continuous and bounded above. We may apply both (iv) and (v) to have

$$\int f dL \leq \liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \limsup \sup \mathbb{E}^*f(X_n) \leq \int f dL$$

Thus

$$\int f dL = \liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) = \limsup_{n \rightarrow +\infty} \mathbb{E}_*f(X_n).$$

In summary, we have proved the Theorem through the following graph. We may check that each point implies all the others by using the right path in :

$$\begin{array}{ccccccc} (i) & \Rightarrow & (vii) & \Rightarrow & (ii) & \Leftrightarrow & (iii) \\ \uparrow & & & & \downarrow & & \uparrow \\ (v) & \Leftrightarrow & (iv) & = & (iv) & (vi) & = & (vi) \end{array}$$

And this shows that the six assertions equivalent.

3. Continuous Mapping Theorem

Let (X_n) , $n \geq 1$, be a sequence of measurable applications with values in the metric space (S, d) converging to the measurable application X with values in (S, d) . Suppose we have a mapping of (S, d) into an other metric space (E, r) . The natural question we may ask ourselves is the following : Does the sequence $(g(X_n))$, $n \geq 1$, weakly converge to $g(X)$?

The answer is easy if g is continuous. To make the ideas clear, denote $Y_n = g(X_n)$, $n \geq 1$, and $Y = g(X)$. Then for any $f \in C_b(E)$, we have $h = (f \circ g) \in C_b(S)$ and for any $n \geq 1$,

$$\mathbb{E}f(Y_n) = \mathbb{E}(f \circ g)(X_n) = \mathbb{E}(h(X_n))$$

and

$$\mathbb{E}f(Y) = \mathbb{E}(f \circ g)(X) = \mathbb{E}(h(X)).$$

Then $\mathbb{E}f(Y_n)$ converges to $\mathbb{E}f(Y)$, whenever (X_n) weakly converges to X . Thus, we may conclude that $(g(X_n))$, $n \geq 1$, weakly converges to $g(X)$.

This result is a particular case of a more general answer given below. Define by D_g the set of all discontinuity points of g . The continuity of g means that D_g is empty. The generalization of the result given below requires that the function g be \mathbb{P}_X -continuous, that is $\mathbb{P}_X(D_g) = 0$. But we cannot write $\mathbb{P}_X(D_g)$ unless we are sure that D_g is measurable. Fortunately, by Lemma 3 in the Appendix Section 7 below, it is a surprising fact that D_g is measurable whatever be g . We have the following more general result

PROPOSITION 6. *Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, B(S))$ be a sequence of measurable applications weakly converging to a measurable application $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (S, B(S))$ (or to the probability measure L) and let g be a mapping of (S, d) into the metric space (E, r) such that g is \mathbb{P}_X -continuous (or L -continuous, then the sequences $g(X_n)$ weakly converges to $g(X)$ (or to $L \circ g^{-1}$).*

Proof. Suppose that $X_n \rightarrow_w L$ with $L(\text{discont}(g)) = 0$. Let F be a closed subset of E . Let us show that the Point (iii) of Portmanteau Theorem 2 holds. Let us first show that,

$$(3.1) \quad \overline{g^{-1}(F)} \subseteq g^{-1}(F) \cup \text{discont}(g).$$

where $\overline{g^{-1}(F)}$ is the closure of $g^{-1}(F)$. Indeed, let $x \in \overline{g^{-1}(F)}$. Then there exists a sequence $(y_n)_{n \geq 1} \in g^{-1}(F)$ such that $y_n \rightarrow x$ and for each $n \geq 1$, $g(y_n) \in F$. From here, we have two cases.

Either $x \in \text{discont}(g)$ and then $x \in g^{-1}(F) \cup \text{discont}(g)$.

Or $x \notin \text{discont}(g)$, that is g is continuous at x . Then, since, $y_n \rightarrow x$, we have $g(y_n) \rightarrow g(x)$. Since the sequence $g(y_n)$ is in F , which is closed, then $g(x) \in F$. This is equivalent to $x \in g^{-1}(F)$ and finally : $x \in g^{-1}(F) \cup \text{discont}(g)$.

We conclude that (3.1) is true by combining both cases.

Now, let us use (3.1) in the following way. We have

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(g(X_n) \in F) = \limsup_{n \rightarrow +\infty} \mathbb{P}_{n \rightarrow +\infty}(X_n \in g^{-1}(F)) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \overline{g^{-1}(F)})$$

and subsequently,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \overline{g^{-1}(F)}) \leq L(\overline{g^{-1}(F)}) \leq L(g^{-1}(F)) + L(\text{discont}(g))$$

This concludes the proof by

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(g(X_n) \in F) \leq L \circ g^{-1}(F).$$

We have the Portmanteau theorem in the general of a metric space case because its proof depends only on general properties of metrics. But when the metric is on the three equivalent ones in \mathbb{R}^k , we may go further and be more precises. The combination of the Portmanteau theorem with the characterization results of probability measures in \mathbb{R}^k leads to stunning and fine results

4. Space \mathbb{R}^k

In this section we focus on the particular metric space $S = \mathbb{R}^k$. Before we begin, let us make some reminder on the characterization of the distributions in \mathbb{R} .

Let $X = \begin{bmatrix} X_1 \\ \cdots \\ X_k \end{bmatrix}$, $X_n = \begin{bmatrix} X_1^{(n)} \\ \cdots \\ X_k^{(n)} \end{bmatrix}$, $n \geq 1$, be random vectors of dimension k .

Terminology. By random vectors in \mathbb{R}^k , we mean measurable applications defined on some measurable space with values in \mathbb{R}^k .

We recall that the probability law of a random vector X of \mathbb{R}^k is characterized by its *distribution function*, defined by

$${}^t(t_1, t_2, \dots, t_k) \mapsto F_X(t_1, t_2, \dots, t_k) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k)$$

or by its characteristic function

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_{j=1}^k i u_j X_j))$$

or by its moment generating function (whenever it exists) defined by

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Psi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_{j=1}^k u_j X_j))$$

or by its probability density function whenever it exists. And it exists respect to the Lebesgues measure for instance if and only if

$${}^t(t_1, t_2, \dots, t_k) \mapsto f_X(t_1, t_2, \dots, t_k) = \frac{\partial^{(k)} F_X(t_1, t_2, \dots, t_k)}{\partial t_1 \partial t_2 \cdots \partial t_k},$$

a.e. with respect to the Lebesgues measures.

It remarkable that these characteristics also play important roles in the theory of weak convergence of random vectors.

We have the following characterizations and criteria.

PROPOSITION 7. *Let $X_n : (\Omega_n, \mathcal{A}_n, P_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$ be a sequence of random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.*

If X_n weakly converges to X , then for any continuity point $t = (t_1, t_2, \dots, t_k)$ of F_X , we have, as $n \rightarrow +\infty$,

$$(4.1) \quad \mathbb{P}_n(X_n \in \prod_{i=1}^k]-\infty, t_i]) \rightarrow F_X(t_1, t_2, \dots, t_k).$$

Proof. Consider the distribution function of X

$$\begin{aligned} F_X(t_1, t_2, \dots, t_k) &= \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k) \\ &= \mathbb{P}(X \in \prod_{i=1}^k]-\infty, t_i]) \end{aligned}$$

Denote $t = (t_1, \dots, t_k)$ and $t(n) = (t_1(n), t_2(n), \dots, t_k(n))$, $n \geq 1$. We have : $t(n) \uparrow t$ (resp $t(n) \downarrow t$) as $n \rightarrow +\infty$ if and only if

$$\forall (1 \leq i \leq k), t_i(n) \uparrow t_i \text{ (resp. } t_i(n) \downarrow t_i \text{) as } n \rightarrow +\infty.$$

Set $A(t) = \prod_{i=1}^k]-\infty, t_i]$. We have as $n \uparrow \infty$,

$$A(t(n)) \downarrow A(t),$$

and by using the Monotone Convergence Theorem,

$$F_X(t) = \mathbb{P}(X \in A(t(n)) \downarrow \mathbb{P}(X \in A(t)) = F_X(t),$$

as $n \rightarrow +\infty$. Then F_X is right continuous at each point t . But

$$A(t(n)) \uparrow A^+(t) = \prod_{i=1}^k]-\infty, t_i[,$$

as $n \rightarrow +\infty$, and next, still by the Monotone Convergence Theorem,

$$F_X(t) = \mathbb{P}(X \in A(t(n)) \uparrow \mathbb{P}(X \in A^+(t)),$$

as $n \rightarrow +\infty$. But we have

$$(4.2) \quad D(t) = A(t) \setminus A^+(t)$$

$$(4.3) \quad = \{x = (x_1, \dots, x_k) \in A(t), \exists 1 \leq i \leq k, x_i = t_i\}.$$

To understand better this formula, let us have a look at it for $k = 1$:

$$]-\infty, a] \setminus]-\infty, a[= \{a\}$$

and for $k = 2$ (a diagram would help) :

$$\begin{aligned} &]-\infty, a] \times]-\infty, b] \setminus]-\infty, a[\times]-\infty, b[\\ & = \{(x, y) \in]-\infty, a] \times]-\infty, b], x = a \text{ or } y = b\} \end{aligned}$$

Hence, if

$$(4.4) \quad \mathbb{P}_\infty(X \in D(t)) = L(D(t)) = 0,$$

we get, as $n \rightarrow \infty$,

$$F_X(t) = \mathbb{P}(X \in A(t(n)) \uparrow \mathbb{P}(X \in A^+(t)) = \mathbb{P}(X \in A(t)) - \mathbb{P}(X \in D(t)) = F_X(t).$$

We conclude that (4.4) is the condition for t to be a continuity point of F_X . Further, $D(t)$ is the boundary of $A(t)$, that is

$$(4.5) \quad \partial A(t) = D(t)$$

To see this, just check that $A(t)$ is closed and that the interior of $A(t)$ is $A^+(t)$. By Point (vi) of Portmanteau Theorem 2, we get that for any continuity point t of F_X ,

$$F_{X_n}(t) = \mathbb{P}(X_n \in A(t)) \rightarrow F_X(t) = \mathbb{P}(X \in A(t)) \text{ as } n \rightarrow +\infty.$$

This ends the proof. Conversely, we will have :

PROPOSITION 8. *Let $X_n : (\Omega_n, \mathcal{A}_n, P_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$ be a sequence of random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, P_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Suppose that for any continuity point t of F_X , we have*

$$(4.6) \quad F_{X_n}(t) = \mathbb{P}(X_n \in A(t)) \rightarrow F_X(t) = \mathbb{P}(X \in A(t)) \text{ as } n \rightarrow +\infty.$$

Then X_n weakly converges to X .

Warning. This proof is lengthy and very technical. It is stated only for people who are training to be a researcher in fundamental mathematics, probability or Statistics. If you are not among this people, you may skip it.

Proof. Suppose that for any $t = (t_1, t_2, \dots, t_k)$ continuity point of F_X and $F_{X_n}(t) \rightarrow F_X(t_1, t_2, \dots, t_k)$, as $n \rightarrow +\infty$.

To show that X_n weakly converges to X , we are going to use Point (ii) of Portmanteau Theorem 2, that is, for any open set G on \mathbb{R}^k , we have

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}(X \in G).$$

Let G be an arbitrary open set in \mathbb{R}^k . By using Proposition , G is a countable union of F_X -continuous intervals in the form

$$G = \bigcup_{j \geq 1}]a^j, b^j],$$

where for all $j \geq 1$, all the points c defined by

$$c_i = a_i^{(j)} \text{ ou } c_i = b_i^{(j)},$$

are continuity points of F_X . In the sequel, \mathcal{U} denotes the set of all bounded F_X -intervals.

Now, by the continuity of the probability measure \mathbb{P}_X , we can find for any $\eta > 0$, an integer m such that

$$(4.7) \quad \mathbb{P}_X(G) - \eta \leq \mathbb{P}_X\left(\bigcup_{j=1}^m]a^j, b^j]\right)$$

We set $A_j =]a^j, b^j]$ and use the Poincaré formula, that is the inclusion-exclusion formula, that gives

$$(4.9) \quad \begin{aligned} \mathbb{P}_X\left(\bigcup_{j=1}^m A_j\right) &= \sum \mathbb{P}_X(A_j) - \sum \mathbb{P}_X(A_i A_j) \\ &+ \sum \mathbb{P}_X(A_i A_j A_k) + \dots + (-1)^{n+1} \mathbb{P}_X(A_1 A_2 \dots A_n) \end{aligned}$$

and

$$(4.11) \quad \begin{aligned} \mathbb{P}_{X_n}\left(\bigcup_{j=1}^m A_j\right) &= \sum \mathbb{P}_{X_n}(A_j) - \sum \mathbb{P}_{X_n}(A_i A_j) \\ &+ \sum \mathbb{P}_{X_n}(A_i A_j A_k) + \dots + (-1)^{n+1} \mathbb{P}_{X_n}(A_1 A_2 \dots A_n) \end{aligned}$$

We are going to handle each of these terms of the expressions above. Let us take one of the terms

$$\mathbb{P}_X(A_{i_1} A_{i_2} \dots A_{i_k}).$$

As showed in Subsection 7.1 in the Annexe Section 7 below, the class \mathcal{U} of F_X -continuous intervals is stable under finite intersection. Thus the $A_{i_1} A_{i_2} \dots A_{i_k}$, which is of the type $]a, b]$, is in \mathcal{U} . It is a F_X -continuous interval. The Lebesgue-Stieltjes Formula, gives

$$\mathbb{P}_X(A_{i_1} A_{i_2} \dots A_{i_k}) = \sum_{\varepsilon \in \{0,1\}^k} (-1)^{(\sum_{1 \leq i \leq k} \varepsilon_i)} F_X(b + \varepsilon * (a - b)).$$

We similarly get that

$$\mathbb{P}_{X_n}(A_{i_1} A_{i_2} \dots A_{i_k}) = \sum_{\varepsilon \in \{0,1\}^k} (-1)^{(\sum_{1 \leq i \leq k} \varepsilon_i)} F_{X_n}(b + \varepsilon * (a - b)).$$

and we are able to apply the assumption of the convergence of F_{X_n} to F_X for continuity points of F_X to have, as $n \rightarrow +\infty$,

$$\mathbb{P}_{X_n}(A_{i_1} A_{i_2} \dots A_{i_k}) \rightarrow \mathbb{P}_X(A_{i_1} A_{i_2} \dots A_{i_k}).$$

By operating term by term in (4.9) and in (4.11), we conclude that, as $n \rightarrow +\infty$,

$$\mathbb{P}_{X_n}\left(\bigcup_{j=1}^m A_j\right) \rightarrow \mathbb{P}_X\left(\bigcup_{j=1}^m A_j\right).$$

Then

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) &= \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \bigcup_{j \geq 1}]a^j, b^j]) \\ &\geq \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \bigcup_{j=1}^m]a^j, b^j]) \geq \mathbb{P}_\infty(X \in G) - \eta, \end{aligned}$$

and this for an arbitrary $\eta > 0$. Then, by letting $\eta \downarrow 0$, we arrive at

$$\liminf \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G),$$

for any open set G in \mathbb{R}^k . We conclude finally that

$$X_n \rightarrow_w X \text{ as } n \rightarrow +\infty.$$

We are moving to characteristic functions. We have the following characterization.

PROPOSITION 9. *Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$ be a sequence of random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ another random vector. Then X_n weakly converges to X as $n \rightarrow +\infty$, if and only if for any point ${}^t(u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,*

$$\Phi_{X_n}(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k) \text{ as } n \rightarrow +\infty.$$

Remark The proof we are proposing here is based of the Stone-Weirstrass Theorem which is an important theorem in spaces of continuous functions defined on a compact set. This theorem is recalled in Proposition 20 in Section 7 below). Another proof, that is more beautiful to us, is provided in Theorem 11 in Chapter 3. This latter is based on the concept of tightness and the Levy continuity theorem.

proof. Recall the definition of the charactersitic function :

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_{j=1}^k i u_j X_j)),$$

which can written as follows.

$${}^t(u_1, u_2, \dots, u_k) \mapsto \exp(\sum_{j=1}^k i u_j X_j) = \cos(\sum_{j=1}^k i u_j X_j) + i \sin(\sum_{j=1}^k i u_j X_j).$$

This is a complex function whose components are bounded and continuous functions of X and by definition, we have

$$\mathbb{E} \exp\left(\sum_{j=1}^k i u_j X_j\right) = \mathbb{E} \cos\left(\sum_{j=1}^k i u_j X_j\right) + i \mathbb{E} \sin\left(\sum_{j=1}^k i u_j X_j\right).$$

Hence, by the very definition of weak convergence, for any point ${}^t(u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,

$$(4.12) \quad \Phi_{X_n}(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k).$$

This proves the direct implication of our proposition. To prove the indirect one, we appeal to the Stone-Weirstrass Theorem (See Proposition 20 in Subsection 7 in Section 7 below).

Set $\Delta(a) = [-a, a]^k$, for $0 < a \in \mathbb{R}$. Define H as the class of bounded functions defined on \mathbb{R}^k by

$$(4.13) \quad {}^t(x_1, x_2, \dots, x_k) \mapsto \sum_{r=1}^m a_r \exp\left(\sum_{j=1}^k \pi n_{j,r} i x_j / a\right)$$

where the coefficients a_r are complex numbers and the $n_{j,r}$ are integer numbers. In other words, the elements of H are finite linear combinations of complex exponential functions of linear function of x_1, x_2, \dots, x_k . Let us notice that each function

$$x_j \mapsto \exp(\pi n_{j,r} i x_j / a)$$

is periodic with period $2a$ so that each function $h \in H$ attains all its values in Δ , that is $\{h(x), x \in \mathbb{R}^k\} \subset \{h(x), x \in \Delta(a)\}$ and then

$$\|h\| = \|h\|_{\Delta(a)}.$$

For each $h \in H$, let us denote by $h_{\Delta(b)}$ the restriction of h on $\Delta(b)$, with $b > 0$, and set $H_b = \{h_{\Delta(b)}, h \in H\}$. We have to check that for each $0 < b < a$, the Stone-Weirstrass Theorem conditions hold for H_b .

(1) H includes all the constants. Let d be any complex number. In (4.13), we take $m = 1, a_1 = d$ and $n_{1,1} = n_{2,1} = \dots = n_{k,1} = 0$. We get $h(x) = a_1 = d$.

(2) H is closed under finite addition and product. This obvious.

(3) Conjugates \bar{h} of elements of h of H remain in H .

(4) Finally H separates the points of \mathbb{R}^k . To see this, take ${}^t(z_1, z_2, \dots, z_k) \neq {}^t(y_1, y_2, \dots, y_k)$ in $\Delta(b)$. Then there exists j_0 such that $z_{j_0} \neq y_{j_0}$. We have $|z_{j_0} - y_{j_0}| \leq 2b < 2a$. Define $h_0 \in H$ by

$$h_0(x) = e^{i\pi x_{j_0}/a} \quad x \in \mathbb{R}^k.$$

Since the equality $e^{i\pi z_{j_0}/a} = e^{i\pi y_{j_0}/a}$ implies that for some $k \in \mathbb{Z}$, $(z_{j_0} - y_{j_0}) = 2ka$, which is impossible for any $k \in \mathbb{Z}$, we have

$$h_0(z) = e^{i\pi z_{j_0}/a} \neq e^{i\pi y_{j_0}/a} = h_0(y).$$

This says that h_0 separates ${}^t(z_1, z_2, \dots, z_k)$ and ${}^t(y_1, y_2, \dots, y_k)$.

The conditions of the Stone-Weirstrass hold. Now Suppose that (4.12) holds and let $f \in C_b(\mathbb{R}^k, \mathbb{R}) \subseteq C_b(\mathbb{R}^k, \mathbb{C})$. Consider f_b the restriction of f on $\Delta(b)$ so that $f_b \in C(\Delta(b), \mathbb{C})$ where $b = b(a) < a$, is an unbounded and increasing function of a ($b(a) = a/2$ for instance). Here, we apply Stone-Weirstrass theorem to find for any $\varepsilon > 0$, a function h in H ,

$$h(x) = \sum_{r=1}^m a_r \exp\left(\sum_j^k i u_{r,j} x_j\right)$$

such that

$$\sup_{x \in \Delta(b)} |f(x) - h(x)| = \|f - h\|_{\Delta(b)} \leq \varepsilon/3.$$

By (4.12), we have

$$\mathbb{E}(h(X_n)) \rightarrow \mathbb{E}(h(X)) \text{ as } n \rightarrow +\infty.$$

Let n_0 such that, for any $n \geq n_0$,

$$(4.14) \quad |\mathbb{E}(h(X_n)) - \mathbb{E}(h(X))| = \left| \int h d\mathbb{P}_n \circ X_n^{-1} - \int h d\mathbb{P} \circ X^{-1} \right| \leq \varepsilon/3$$

We have

$$\begin{aligned} \mathbb{E}(f(X_n)) - \mathbb{E}(f(X)) &= \left(\int f d\mathbb{P}_n \circ X_n^{-1} - \int h d\mathbb{P}_n \circ X_n^{-1} \right) \\ &\quad + \left(\int h d\mathbb{P}_n \circ X_n^{-1} - \int_c h d\mathbb{P} \circ X^{-1} \right) \\ &\quad + \left(\int_c h d\mathbb{P} \circ X^{-1} - \int_c f d\mathbb{P} \circ X^{-1} \right). \end{aligned}$$

The first term satisfies

$$\begin{aligned}
\mathbb{E} \left| \int f \, d\mathbb{P}_n \circ X_n^{-1} - \int h \, d\mathbb{P}_n \circ X_n^{-1} \right| &\leq \int_{\Delta(b)} |f - h| \, d\mathbb{P}_n \circ X_n^{-1} \\
&\quad + \int_{\Delta^c(b)} |f - h| \, d\mathbb{P}_n \circ X_n^{-1} \\
(4.15) \qquad \qquad \qquad &\leq \varepsilon/3 + (\|f\| + \|h\|)\mathbb{P}_n(X_n \in \Delta^c(b)).
\end{aligned}$$

By treating the third term in the same manner, we also get

$$(4.16) \quad \mathbb{E} \left| \int f \, d\mathbb{P}_\infty \circ X^{-1} - \int h \, d\mathbb{P}_\infty \circ X^{-1} \right| \leq \varepsilon/3 + (\|f\| + \|h\|) \mathbb{P}_\infty(X \in \Delta^c(b))$$

By putting together Formulas, (4.14), (4.15) and (4.16), we get for each fixed $n \geq n_0$,

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq \varepsilon + (\|f\| + \|h\|)(\mathbb{P}_n(X_n \in \Delta^c(b)) + \mathbb{P}_\infty(X \in \Delta^c(b))).$$

For each fixed $n \geq n_0$, by letting $a \uparrow +\infty$, we have $b(a) \uparrow +\infty$ and next, $\mathbb{P}_n(X_n \in \Delta^c(b)) + \mathbb{P}_\infty(X \in \Delta^c(b)) \downarrow 0$. Then for each $n \geq n_0$

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq \varepsilon$$

This shows that

$$\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X)) \text{ as } n \rightarrow +\infty$$

and then

$$X_n \rightsquigarrow X \text{ as } n \rightarrow +\infty.$$

By putting together (7), (8) and (9), we have the full Pormanteau Theorem in \mathbb{R}^k .

THEOREM 3. *Let k be a positive integer. The sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, ≥ 1 , weakly converges to the random vector $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ if and only if one of these assertions holds.*

(i) *For any real-valued continuous and bounded function f defined on \mathbb{R}^k ,*

$$\lim_{n \rightarrow +\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

(ii) For any open set G in \mathbb{R}^k ,

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G).$$

(iii) For any closed set G of S , we have

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \leq \mathbb{P}_\infty(X \in G).$$

(iv) For any inferior semi-continuous and bounded below function f , we have

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(v) For any superior semi-continuous and bounded above function f , we have

$$\limsup_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \mathbb{E}f(X).$$

(vi) For any Borel set B of S that is \mathbb{P}_X -continuous, that is $\mathbb{P}_\infty(X \in \partial B) = 0$, we have

$$\lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \mathbb{P}_X(B) = \mathbb{P}_\infty(X \in B).$$

(vii) For any nonnegative and bounded Lipschitz function f , we have

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(viii) For any continuity point $t = (t_1, t_2, \dots, t_k)$ of F_X , we have,

$$F_{X_n}(t) \rightarrow F_X(t) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, F_{X_n} is the distribution function of X_n and F_X that of X .

(ix) For any point $u = (u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \rightarrow \Phi_X(u) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, Φ_{X_n} is the characteristic function of X_n and Φ_X is that of X .

By putting together (7), (8) and (9), we have the full Portmanteau Theorem in \mathbb{R}^k .

THEOREM 4. *Let k be a positive integer. The sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, weakly converges to the random vector $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ if and only if one of these assertions holds.*

(i) *For any real-valued continuous and bounded function f defined on \mathbb{R}^k ,*

$$\lim_{n \rightarrow +\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

(ii) *For any open set G in \mathbb{R}^k ,*

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G).$$

(iii) *For any closed set G of S , we have*

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \leq \mathbb{P}_\infty(X \in G).$$

(iv) *For any inferior semi-continuous and bounded below function f , we have*

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(v) *For any superior semi-continuous and bounded above function f , we have*

$$\limsup_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \mathbb{E}f(X).$$

(vi) *For any Borel set B of S that is \mathbb{P}_X -continuous, that is $\mathbb{P}_\infty(X \in \partial B) = 0$, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \mathbb{P}_X(B) = \mathbb{P}_\infty(X \in B).$$

(vii) *For any nonnegative and bounded Lipschitz function f , we have*

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(viii) *For any continuity point $t = (t_1, t_2, \dots, t_k)$ of F_X , we have,*

$$F_{X_n}(t) \rightarrow F_X(t) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, F_{X_n} is the distribution function of X_n and F_X that of X .

(ix) For any point $u = (u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \mapsto \Phi_X(u) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, Φ_{X_n} is the characteristic function of X_n and Φ_X is that of X

The characteristic function as a tool of weak convergence is also used through the following criteria.

Wold Criterion. The sequence $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ weakly converges to $X \in \mathbb{R}^k$, as $n \rightarrow +\infty$ if and only if for any $a \in \mathbb{R}^k$, the sequence $\{< a, X_n >, n \geq 1\} \subset \mathbb{R}$ weakly converges to $X \in \mathbb{R}$ as $n \rightarrow +\infty$.

Proof. The proof is quick and uses the notation above. Suppose that X_n weakly converges to X in \mathbb{R}^k as $n \rightarrow +\infty$. By using the convergence of characteristic functions, we have for any $u \in \mathbb{R}^k$

$$\mathbb{E}(\exp(i < X_n, u >)) \rightarrow \mathbb{E}(\exp(i < X, u >)) \text{ as } n \rightarrow +\infty.$$

It follows for any $a \in \mathbb{R}^k$ and for any $t \in \mathbb{R}$, we have

$$(4.17) \quad \mathbb{E}(\exp(it < X_n, a >)) \rightarrow \mathbb{E}(\exp(it < X, a >)) \text{ as } n \rightarrow +\infty.$$

that is, by taking $u = ta$ in the formula above, and by denoting $Z_n = < X_n, a >$ and $Z = < X, a >$

$$\mathbb{E}(\exp(itZ_n)) \rightarrow \mathbb{E}(\exp(itZ)) \text{ as } n \rightarrow +\infty.$$

This means that $Z_n \rightsquigarrow Z$, that is $< a, X_n >$ weakly converges to $< a, X >$.

Conversely, suppose that for any $a \in \mathbb{R}^k$, the sequence $\{< a, X_n >, n \geq 1\} \subset \mathbb{R}$ weakly converges to $X \in \mathbb{R}$ as $n \rightarrow +\infty$. Then by taking $t = 1$ in (4.17) we get for any $a = u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i < X, u >)) \rightarrow \mathbb{E}(\exp(i < X, u >)) \text{ as } n \rightarrow +\infty.$$

which means that $X_n \rightsquigarrow X$ as $n \rightarrow +\infty$.

5. Theorem of Scheffé

In the preceding section, we linked the weak convergence to some characteristics of random vectors distributions, in particular the distribution functions and the characteristic functions. Now, what happens for the probability density functions? The theorem of Scheffé goes

beyond the particular case of \mathbb{R}^k and gives a very general answer as follows.

THEOREM 5. . *Let λ be a une mesure on some measurable space (E, \mathcal{B}) . Let $p, (p_n)_{n \geq 1}$ be probability densities with respect to λ , that real valued, nonnegative and measurable functions defined on E such that*

$$(5.1) \quad \forall n \geq 1, \int p_n d\lambda = \int p d\lambda = 1.$$

Suppose that

$$p_n \rightarrow p, \quad \lambda - a.e.$$

Then

$$(5.2) \quad \sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda \rightarrow 0$$

Proof. Suppose $p_n \rightarrow p, \lambda - pp$. Set $\Delta_n = p - p_n$. Then (5.1) implies

$$\int \Delta_n d\lambda = 0.$$

Then, for $B \in \mathcal{B}$,

$$\int_{B^c} \Delta_n d\lambda = \int \Delta_n d\lambda - \int_B \Delta_n d\lambda = - \int_B \Delta_n d\lambda.$$

Thus,

$$(5.3) \quad 2 \left| \int_B \Delta_n d\lambda \right| = \left| \int_B \Delta_n d\lambda \right| + \left| \int_{B^c} \Delta_n d\lambda \right|$$

$$(5.4) \quad \leq \int_B |\Delta_n| d\lambda + \int_{B^c} |\Delta_n| d\lambda \leq \int |\Delta_n| d\lambda,$$

meaning that

$$(5.5) \quad \left| \int_B \Delta_n d\lambda \right| \leq \frac{1}{2} \int |\Delta_n| d\lambda.$$

By taking $B = (\Delta_n \geq 0)$ in (5.4), we get

$$2 \left| \int_B \Delta_n d\lambda \right| = \left| \int_B \Delta_n^+ d\lambda \right| + \left| \int_{B^c} -\Delta_n^- d\lambda \right| = \int \Delta_n^+ d\lambda + \int \Delta_n^- d\lambda = \int |\Delta_n| d\lambda.$$

By putting together the two last formulas, we have

$$(5.6) \quad \sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda.$$

Now we get,

$$0 \leq \Delta_n^+ = \max(0, p - p_n) \leq p.$$

Besides, we have

$$\int \Delta_n^+ d\lambda = \int_{(\Delta_n \geq 0)} \Delta_n d\lambda = \int \Delta_n d\lambda - \int_{(\Delta_n \leq 0)} \Delta_n d\lambda = \int_{(\Delta_n \leq 0)} -\Delta_n d\lambda = \int \Delta_n^- d\lambda,$$

so that

$$(5.7) \quad \int |\Delta_n| d\lambda = 2 \int \Delta_n^+ d\lambda$$

Here, we apply the Fatou-Lebesgues Dominated Theorem to

$$0 \leq \Delta_n^+ \leq |\Delta_n| \rightarrow 0 \text{ } \lambda - pp \text{ as } n \rightarrow +\infty, \text{ and } 0 \leq \Delta_n^+ \leq p$$

We get

$$\int \Delta_n^+ d\lambda \rightarrow 0,$$

in virtue of (5.6),

$$\sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda = \int \Delta_n^+ d\lambda \rightarrow 0.$$

which puts and end to the proof.

The Theorem of Scheffé may be applied to probability densities in \mathbb{R}^k with respect to the Lebesgue measure or to a counting measure.

PROPOSITION 10. *These two assertions hold.*

(A) Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ be random vectors and $X : \Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ another random vector, all of them absolutely continuous with respect to the Lebesgues measure denoted as λ_k . Denote f_{X_n} the probability density function of X_n , $n \geq 1$ and by f_X the probability density function of X . Suppose that we have

$$f_{X_n} \rightarrow f_X, \quad \lambda_k - a.e., \quad \text{as } n \rightarrow +\infty.$$

Then X_n weakly converges to X as $n \rightarrow +\infty$.

(B) Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ be discrete random vectors and $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ another discrete random vector. For each n , define D_n the countable support of X_n , that

$$\mathbb{P}_n(X_n \in D_n) = 1 \text{ and for each } x \in D_n, \mathbb{P}(X_n = x) \neq 0.$$

and D_∞ the countable support of X . Set $D = D_\infty \cup (\cup_{n \geq 1} D_n)$ and denote ν as the counting measure on D . Then the probability densities of the X_n and of X with respect to ν are defined on D by

$$f_{X_n}(x) = \mathbb{P}_n(X_n = x), \quad n \geq 1, \quad f_X(x) = \mathbb{P}_\infty(X = x), \quad x \in D.$$

If

$$(\forall x \in D), f_{X_n}(x) \rightarrow f_X(x).$$

Then X_n weakly converges to X .

6. Weak Convergence and Convergence in Probability on one Probability Space

In this section, we place the weak convergence limit in the general frame of the convergence of random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a metric space (S, d) . We already saw that the weak convergence of random variables does not require from them and from the weak limit random variable that they are defined on a common probability space. In the particular case where this happens, and only in this case, we are able to have interesting relations with other types of convergences.

Conversely, the powerful theorem of Skorohod-Wichura-Dudley allows to transform any weak convergence, under specific conditions on the space S , to an almost-sure convergence of versions of the sequences and on the limit. In this text, this theorem is only proved when S is the real line \mathbb{R} in Chapter 4. The proof is expected in a more general book on weak convergence.

Let us begin with the definitions.

6.1. Definitions. In all this section, except in the Subsection 12, the random variables $(X_n)_{n \geq 0}$, $(Y_n)_{n \geq 0}$, etc., and the random variables X , Y , etc. are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and have their values in the metric space (S, d) . We will also have to use constants c in S . So you will not find probability measures \mathbb{P}_∞ , and \mathbb{P}_n , $n \geq 1$, here.

(a) Almost-sure convergence.

The sequence $(X_n)_{n \geq 0}$ converges almost-surely to X as $n \rightarrow \infty$, denoted as $X_n \rightarrow X$, a.s. as $n \rightarrow \infty$, if and only if the subspace of Ω on which $(X_n)_{n \geq 0}$ fails to converge to X is a \mathbb{P} -null set, that is

$$\mathbb{P}(\{\omega \in \Omega, X_n \not\rightarrow X\}) = \mathbb{P}(\{\omega \in \Omega, d(X_n, X) \not\rightarrow 0\}) = 0.$$

This may be expressed as

$$(X_n \not\rightarrow X) = \bigcup_{k \geq 1} \bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}).$$

and this is surely measurable because of the continuity of the metric d . This leads to the new definition : $(X_n)_{n \geq 0}$ almost-surely converges to X as $n \rightarrow \infty$, if and only if :

$$(6.1) \quad \forall k \geq 1, \mathbb{P} \left(\bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) \right) = 0.$$

(b) Convergence in probability.

The sequence $(X_n)_{n \geq 0}$ converges in probability to X , as $n \rightarrow +\infty$, denoted as $X_n \rightarrow_{\mathbb{P}} X$, if and if

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

Now, we are going to make a brief comparaison between these two types of convergence. The following proposition is already knew to the reader in the case where S is \mathbb{R} .

PROPOSITION 11. *If $X_n \rightarrow X$ a.s. as $n \rightarrow +\infty$, then $X_n \rightarrow_{\mathbb{P}} X$ as $n \rightarrow +\infty$*

Proof. The proof is the same as in \mathbb{R} . Suppose that $X_n \rightarrow X$ a.s. as $n \rightarrow +\infty$. We have to prove (6.1). We have for $k \geq 1$,

$$(d(X_n, X) > k^{-1}) \subset \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) =: B_{n,k}.$$

But the sequence $B_{n,k}$ is non-decreasing in n to

$$\bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) =: B_k$$

and for any $n \geq 0$ and $k \geq 1$,

$$(6.2) \quad \mathbb{P}(d(X_n, X) > k^{-1}) \leq \mathbb{P}(B_{n,k}).$$

By the continuity of the probability,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, X) > k^{-1}) \leq \lim_{n \rightarrow \infty} \mathbb{P}(B_{n,k}) = \mathbb{P}(B_k) = 0,$$

where we applied (6.1) to the left member of (6.2).

Now we are going to give a number of relations between the convergence in probability and the weak convergence

6.2. Weak Convergence and Convergence in Probability.

Before we step in the comparison results, we have to enrich the Portmanteau Theorem 2 by this supplementary point.

LEMMA 1. *The sequence $(X_n)_{n \geq 0}$ weakly converges to X as $n \rightarrow +\infty$ if and only if*

(viii) *For any bounded Lipschitz function $f : S \rightarrow \mathbb{R}$*

$$\mathbb{E}f(X_n) \rightarrow f(X).$$

as $n \rightarrow +\infty$.

Proof. Let us place ourselves in the proof of Portmanteau Theorem 2. Now (vii) is a subcase of (viii), and then (viii) \implies (vii). Now if (vii) holds, we may take the infimum A and the supremum B of a bounded Lipschitz function f . By applying Point (vii) to $f - A$ and to $-f + B$, we get (viii). Then we have (vii) \iff (viii).

We are going to state a number of properties.

In the sequence, all limits in presence of subscripts n are meant as $n \rightarrow +\infty$ unless the contrary is specified.

(a) The convergence in probability implies the weak convergence

PROPOSITION 12. *If $X_n \xrightarrow{\mathbb{P}} X$ as $n \rightarrow +\infty$, then $X_n \rightsquigarrow X$ as $n \rightarrow +\infty$.*

Proof. Suppose that $X_n \xrightarrow{\mathbb{P}} X$. Let us show that $X_n \rightsquigarrow X$ by using Point (viii) of Lemma 1 above. Let f be a Lipschitz bounded function of coefficient $\ell > 0$ and of bound M . We have for any $n \geq 0$,

$$|f(X_n) - f(X)| \leq \ell d(X_n, X).$$

We have for any $n \geq 0$ and for any $\varepsilon > 0$,

$$\begin{aligned} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| &\leq \mathbb{E}|f(X_n) - f(X)| \\ &\leq \int_{(d(X_n, X) \leq \varepsilon)} |f(X_n) - f(X)| d\mathbb{P} \\ &\quad + \int_{(d(X_n, X) > \varepsilon)} |f(X_n) - f(X)| d\mathbb{P}. \end{aligned}$$

But for any $n \geq 0$ and for any $\varepsilon > 0$,

$$\int_{(d(X_n, X) \leq \varepsilon)} \ell d(X_n, X) d\mathbb{P} \leq \ell \varepsilon.$$

Furthermore, for any $n \geq 0$ and for $\varepsilon > 0$,

$$\int_{(d(X_n, X) > \varepsilon)} |f(X_n) - f(X)| d\mathbb{P} \leq \int_{(d(X_n, X) > \varepsilon)} 2M d\mathbb{P} \leq 2M \mathbb{P}(d(X_n, X) > \varepsilon).$$

Then for any $n \geq 0$ and for any $\varepsilon > 0$,

$$|\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq \ell \varepsilon + 2M \mathbb{P}(d(X_n, X) > \varepsilon).$$

Then for any $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq \ell \varepsilon.$$

By letting $\varepsilon \downarrow 0$, we get

$$\mathbb{E}f(X_n) \longrightarrow \mathbb{E}f(X),$$

which finishes the proof.

(b) Weak convergence and convergence in probability to a constant are equivalent.

PROPOSITION 13. *We have the following equivalence : $X_n \rightarrow_{\mathbb{P}} c$ as $n \rightarrow +\infty$ if and only if $X_n \rightsquigarrow c$ as $n \rightarrow +\infty$.*

Proof. The implication $(X_n \rightarrow_{\mathbb{P}} c) \Rightarrow (X_n \rightsquigarrow c)$ comes from Proposition 12. Let us prove that $(X_n \rightsquigarrow c) \Rightarrow (X_n \rightarrow_{\mathbb{P}} c)$. Suppose that $(X_n \rightsquigarrow c)$. Let $\varepsilon > 0$. Point (ii) of Portmanteau Theorem 2 gives

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) < \varepsilon) &= \liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in B(c, \varepsilon)) \leq \mathbb{P}(c \in B(c, \varepsilon)) \\ &= \mathbb{P}(d(c, c) > \varepsilon) \\ &= \mathbb{P}(\Omega) = 1. \end{aligned}$$

Then

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) \geq \varepsilon) = 1 - \liminf_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) < \varepsilon) \leq 1 - 1 = 0.$$

Then for any $\varepsilon > 0$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) > \varepsilon) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) \geq \varepsilon) = 0.$$

Hence $X_n \rightarrow_{\mathbb{P}} c$.

(c) Two equivalent sequences in probability weakly converge to the same limit if one of them does..

PROPOSITION 14. *If $X_n \rightsquigarrow X$ and $d(X_n, Y_n) \rightarrow_{\mathbb{P}} 0$ as $n \rightarrow +\infty$, then $Y_n \rightsquigarrow X$.*

Proof. Suppose $X_n \rightsquigarrow X$ and $d(X_n, Y_n) \rightarrow_{\mathbb{P}} 0$. Let us prove that $Y_n \rightsquigarrow X$ by using Point (vii) of Lemma 12 above. Let f be a bounded Lipschitz with coefficient $\ell > 0$ and bound M . We have for any $n \geq 0$ and $\varepsilon > 0$,

$$\begin{aligned} |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| &\leq \mathbb{E}|f(Y_n) - f(X)| \\ &\leq \mathbb{E}|f(X_n) - f(X)| + \mathbb{E}|f(Y_n) - f(X_n)|. \end{aligned}$$

By applying Point (vii) of Lemma 1 above and by using the weak limit $X_n \rightsquigarrow X$, we get

$$\limsup_{n \rightarrow +\infty} |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| \leq \limsup_{n \rightarrow +\infty} \mathbb{E}|f(Y_n) - f(X_n)|.$$

Now we use the same method used in the proof of Proposition 12 to have

$$\begin{aligned} \mathbb{E} |f(Y_n) - f(X_n)| &\leq \int_{(d(Y_n, X_n) \leq \varepsilon)} |f(Y_n) - f(X_n)| d\mathbb{P} \\ &\quad + \int_{(d(Y_n, X_n) > \varepsilon)} |f(Y_n) - f(X_n)| d\mathbb{P} \\ &\leq \ell\varepsilon + 2M d(Y_n, X_n), \end{aligned}$$

which tends to zero as $n \rightarrow +\infty$ and next $\varepsilon \downarrow 0$. We conclude that

$$\lim_{n \rightarrow +\infty} \sup |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| = 0.$$

(d) Slutsky's Theorem.

We have the following important and yet simple tool in weak convergence.

PROPOSITION 15. *If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $(X_n, Y_n) \rightsquigarrow (X, c)$*

Proof. Let $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{\mathbb{P}} c$. We want to show that $(X_n, Y_n) \rightsquigarrow (X, c)$. We first remark that $Y_n \xrightarrow{\mathbb{P}} c$ since $Y_n \rightsquigarrow c$. Next, on S^2 endowed with the euclidian metric,

$$d_e((x', y'), (x'', y'')) = \sqrt{d(x', x'')^2 + d(y', y'')^2},$$

we have

$$d_e((X_n, Y_n), (X_n, c)) = d(Y_n, c).$$

It comes that for any $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \sup \mathbb{P}(d_e((X_n, Y_n), (X_n, c)) > \varepsilon) = \lim_{n \rightarrow +\infty} \sup \mathbb{P}(d(Y_n, c) > \varepsilon) = 0,$$

since $Y_n \xrightarrow{\mathbb{P}} c$. Then $d_e((X_n, Y_n), (X_n, c)) \xrightarrow{\mathbb{P}} 0$. By Proposition 14, it is enough to have the weak limit (X_n, c) which will be that of (X_n, Y_n) .

To show the weak convergence of (X_n, c) to (X, c) , we consider a real bounded and continuous function $g(\cdot, \cdot)$ defined on S^2 and try show that $\mathbb{E}g(X_n, c) \rightarrow \mathbb{E}g(X, c)$. But it comes from that c is fixed and the function $f(x) = g(x, c)$ is bounded and continuous and then $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$ since $X_n \rightsquigarrow X$. But $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$ is $\mathbb{E}g(X_n, c) \rightarrow \mathbb{E}g(X, c)$. This finishes the proof.

(d) Coordinatewise convergence in probability.

PROPOSITION 16. *Let $X_n \rightarrow_{\mathbb{P}} X$ and $Y_n \rightarrow_{\mathbb{P}} Y$ if and only if $(X_n, Y_n) \rightarrow_{\mathbb{P}} (X, Y)$.*

Proof. Suppose that $X_n \rightarrow_{\mathbb{P}} X$ and $Y_n \rightarrow_{\mathbb{P}} Y$. Let us use the Manhattan distance on S^2 :

$$d_m((x', y')x', y'), (x'', y'')) = d(x', x'') + d(y', y'').$$

For any $\varepsilon > 0$, $\limsup_{n \rightarrow +\infty} \mathbb{P}(d_m((X_n, Y_n), (X, Y)) > \varepsilon)$ is

$$\begin{aligned} &= \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, Y_n) + d(X, Y) > \varepsilon) \\ &\leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon/2) + \mathbb{P}(d(Y_n, Y) > \varepsilon/2) \\ &\leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon/2) + \limsup_{n \rightarrow +\infty} \mathbb{P}(d(Y_n, Y) > \varepsilon/2) \\ &= 0. \end{aligned}$$

Conversely, suppose that $(X_n, Y_n) \rightarrow_{\mathbb{P}} (X, Y)$. Then for any $n \geq 1$,

$$d(X_n, X) \leq d(X_n, X) + d(Y_n, Y) = d_m((X_n, X), (Y_n, Y)) \rightarrow_{\mathbb{P}} 0 \text{ as } n \rightarrow +\infty.$$

Then $d(X_n, X) \rightarrow_{\mathbb{P}} 0$ as $n \rightarrow +\infty$ and, in the same manner, $d(Y_n, Y) \rightarrow_{\mathbb{P}} 0$.

6.3. Skorohod-Wichura Theorem. We only state this result in a complete and separable metric space.

THEOREM 6. *Let $(X_n)_{n \geq 0}$, $n \geq 1$ and X be of measurable applications with values in (S, d) , a complete and separable space, not necessarily defined on the same probability space.*

If $X_n \rightsquigarrow X$, then there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding measurable applications $(Y_n)_{n \geq 0}$ and Y such that

$$\mathbb{P}_X = \mathbb{P}_Y \text{ and } (\forall n \geq 0, \mathbb{P}_{X_n} = \mathbb{P}_{Y_n})$$

and

$$Y_n \rightarrow Y, \text{ p.s.}$$

This theorem is powerful and may reveal itself very usefull in a great number of situations. You will find a proof of it for $S = \mathbb{R}$ in Chapter 4, Theorem ??.

7. Appendix

7.1. F -continuous intervals, where F is a distribution function. Let \mathbb{P} be a probability measure \mathbb{P} on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Consider its distribution function

$$(x_1, \dots, x_k) \mapsto F(x_1, \dots, x_k) = P\left(\prod_{i=1}^k]-\infty, x_i]\right).$$

7.1.1. *F -continuous intervals.* Let

$$]a, b] = \prod_{i=1}^k]a_i, b_i].$$

be and interval of \mathbb{R}^k . Define

$$E(a, b) = \{c = (c_1, \dots, c_k) \in \mathbb{R}^k, \forall 1 \leq i \leq k, (c_i = a_i \text{ ou } c_i = b_i)\}.$$

We may use extra-notations to get compacts forms of $E(a, b)$. Define the product of k -tuples term by term

$$(x_1, \dots, x_k) * (y_1, \dots, y_k) = (x_1 y_1, \dots, x_k y_k).$$

We also have

$$E(a, b) = \{b + \varepsilon * (a - b), \varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{0, 1\}^k\}.$$

We say that the interval (a, b) is F -continuous if and only if (a, b) is bounded and each element of $E(a, b)$ is a continuity point of F , that is

$$\forall c \in E(a, b), \mathbb{P}(\partial] - \infty, c]) = 0.$$

Let \mathcal{U} be the class of all F -continuous intervals. By convention, we say that the empty set is an F -continuous interval. Here are some properties of \mathcal{U} .

7.1.2. \mathcal{U} is stable by finite intersection. Take $]a, b] = \prod_{i=1}^k]a_i, b_i] \in \mathcal{U}$ and $]c, d] = \prod_{i=1}^k]c_i, d_i] \in \mathcal{U}$. We have

$$]a, b] \cap]c, d] = \prod_{i=1}^k]a_i \vee c_i, b_i \wedge d_i] =]\alpha, \beta]$$

where $x \vee y$ and $x \wedge y$ respectively stand for the maximum and the minimum of x and y , and $\alpha = (a_1 \vee c_1, \dots, a_k \vee c_k)$ and $\beta = (b_1 \wedge d_1, \dots, b_k \wedge d_k)$. If $]a, b] \cap]c, d]$ is empty, it is in \mathcal{U} . Otherwise, none of the factor $]a_i \vee c_i, b_i \wedge d_i]$ is empty. We are going to show that :

$$(7.1) \quad \forall e \in E(\alpha, \beta), \partial] - \infty, e] \subset \bigcup_{z \in E(a, b) \cup E(c, d)} \partial] - \infty, z].$$

Indeed, take $e \in E(\alpha, \beta)$. We have

$$e_i = a_i \vee c_i \text{ ou } b_i \wedge d_i, \quad 1 \leq i \leq k.$$

Take $t \in \partial] - \infty, e]$. This means that

$$(t_i \leq e_i, \quad 1 \leq i \leq k) \text{ and } (\exists i_0, t_{i_0} = e_{i_0})$$

Since $] \alpha, \beta]$ is included in $]a, b]$ and in $]c, d]$, t satisfies

$$t_i \leq b_i \text{ et } t_i \leq d_i, \quad 1 \leq i \leq k.$$

Now, let us consider i_0 such such $t_{i_0} = e_{i_0}$. We have four cases

$$\left\{ \begin{array}{l} t_{i_0} = e_{i_0} = a_{i_0} \vee c_{i_0} = a_{i_0} \implies t_{i_0} = a_{i_0} \text{ et } t_i \leq b_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = a_{i_0} \vee c_{i_0} = c_{i_0} \implies t_{i_0} = c_{i_0} \text{ et } t_i \leq d_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = b_{i_0} \wedge d_{i_0} = b_{i_0} \implies t_{i_0} = b_{i_0} \text{ et } t_i \leq b_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = b_{i_0} \wedge d_{i_0} = d_{i_0} \implies t_{i_0} = d_{i_0} \text{ et } t_i \leq d_i, 1 \leq i \leq k \end{array} \right.$$

We are going to conclude by considering each line of the formula above.

First line : $t \in \partial] - \infty, z_1]$ where $z_1 = (b_1, \dots, b_{i_0-1}, a_{i_0}, b_{i_0+1}, b_k) \in E(a, b)$.

Second line : $t \in \partial] - \infty, z_2]$ where $z_2 = (d_1, \dots, d_{i_0-1}, c_{i_0}, d_{i_0+1}, d_k) \in E(c, d)$.

Third line : $t \in \partial] - \infty, b]$ and of course $b \in E(a, b)$.

Fourth line : $t \in \partial] - \infty, d]$ and of course $d \in E(c, d)$. So t is one of the $\partial] - \infty, z]$ with $z \in E(a, b) \cup E(c, d)$.

So 7.1 holds, and since the union is a finite union of null sets, we have

$$\forall e \in E(\alpha, \beta), P(\partial] - \infty, e]) = 0.$$

Therefore, \mathcal{U} is stable by finite intersection.

LEMMA 2. *Every neighborhood of an arbitrary point x includes a F -continuous interval $]a, b]$ containing x .*

Let V be a neighborhood of x . There exists an interval such that $]a, b[$

$$x \in \prod_{i=1}^k]a_i, b_i[.$$

Set

$$\varepsilon_0 = \min(x_i - a_i, 1 \leq i \leq k) \wedge \min(b_i - x_i, 1 \leq i \leq k),$$

denote by $\delta = (1, \dots, 1)$ the vector of \mathbb{R}^k whos all components are equal to one. We have for $0 < \varepsilon < \varepsilon_0$,

$$]a + \varepsilon\delta, x + \varepsilon\delta] \subset]a, b[.$$

Each point e of $E(a + \varepsilon\delta, x + \varepsilon\delta)$ is of the form

$$t(\varepsilon) = (t_1 + \varepsilon, t_2 + \varepsilon, \dots, t_k + \varepsilon)$$

with, of course, $t_i = a_i$ ou $t_i = x_i$. For any choice of these $t = (t_1, \dots, t_k)$, the sets $\partial] - \infty, t(\varepsilon)]$ are disjoint. Then, by Proposition 19 below we have

$$\mathbb{P}(\partial] - \infty, t(\varepsilon)]) > 0,$$

except, eventually, when ε is out of countable set $D_t \subset]0, \varepsilon_0[$. But $D = \cup_t D(t) \subset]0, \varepsilon_0[$ is countable, since it is at most, a union of 2^k countable sets. The surely, we may pick a value of ε out of $]0, \varepsilon_0[$, such that for any vector e satisfying

$$e_i = a_i + \varepsilon \text{ or } x_i + \varepsilon$$

we have

$$P(\partial] - \infty, t(\varepsilon)]) = 0$$

and

$$x \in]a + \varepsilon\delta, x + \varepsilon\delta] \subset]a, b[.$$

We just proved that there exists $]A_x, B_x[=]a + \varepsilon\delta, x + \varepsilon\delta/2[$ and $]a_x, a_x] =]a + \varepsilon\delta, x + \varepsilon\delta]$ such that

$$(7.2) \quad x \in]A_x, B_x[\subset]a_x, b_x] \subset V.$$

Let us use this to show that *any open set G of \mathbb{R}^k is a countable union of F -continuous intervals.*

Indeed, by (7.2), any open set G may be written as

$$G = \bigcup_{x \in G}]A_x, B_x[.$$

Since \mathbb{R}^k is a separable space, this open cover reduces to a countable cover, that is there exists a sequence $(x_j)_{j \geq 0} \subset G$ such that

$$G = \bigcup_{j \geq 0}]A_{x_j}, B_{x_j}[.$$

We finally get

$$G = \bigcup_{j \geq 0}]a_{x_j}, b_{x_j}].$$

where the $]a_{x_j}, b_{x_j}]$ are F -continuous intervals. We have this proposition.

PROPOSITION 17. *Let F be any probability distribution function on \mathbb{R}^k , $k \geq 1$. Then any open G set in \mathbb{R}^k is a countable union of F -continuous intervals of the form $]a, b]$ or $]a, b[$, where by definition, an interval (a, b) is F -continuous if and only if, for any*

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k) \in \{0, 1\}^k$$

,
the point

$$b + \varepsilon * (a - b) = (b_1 + \varepsilon_1(a_1 - b_1), b_2 + \varepsilon_2(a_2 - b_2), \dots, b_k + \varepsilon_k(a_k - b_k))$$

is a continuity point of F .

7.2. Fonctions semi-continues. A function $f : S \mapsto \overline{\mathbb{R}}$, where S is a metric space, is continuous if and only if

(i) For any $x \in \mathbb{R}$, for any $\varepsilon > 0$, there exists a neighborhood V of x such that

$$y \in V \Rightarrow f(y) \in]f(x) - \varepsilon, f(x) + \varepsilon[.$$

In this formula, we use the whole interval $]f(x) - \varepsilon, f(x) + \varepsilon[$ in the definition. But we might be interested only by one the half intervals. This gives semi-continuous functions. Precisely, a real-valued function f is superior semi-continuous (s.s.c for short) if and only if

(ii) For any $x \in \mathbb{R}$, for any $\varepsilon > 0$, there exists a neighborhood V of x such that

$$y \in V \Rightarrow f(y) < f(x) + \varepsilon,$$

and a real-valued function f is inferior semi-continuous (i.s.c for short) if and only if

(iii) For any $x \in \mathbb{R}$, for any $\varepsilon > 0$, there exists a neighborhood V of x such that

$$y \in V \Rightarrow f(y) > f(x) - \varepsilon.$$

We have two immediate remarks.

(a) A real function is continuous if and only if it is both *s.s.c* and *i.s.c*.

(b) A real function f is **s.s.c.** if and only if its opposite $-f$ is *i.s.c*.

Here is characterization for real valued semi-continuous functions.

PROPOSITION 18. *We have the following properties :*

(1) *Let $f : S \mapsto \overline{\mathbb{R}}$ a superior semi-continuous function if and only if the set $(f \geq c)$ is closed for any real number $c \in \mathbb{R}$.*

(2) *If f is inferior semi-continuous if and only if the set $(f \leq c)$ is closed for any real number $c \in \mathbb{R}$.*

(3) *If f is s.s.c or i.s.c, it is measurable.*

Proof. Proof of Point (1). Let us begin by the direct implication. Let f be a *s.s.c* function from S to \mathbb{R} . Let us show that the set $(f \geq c)$ is closed by showing that the set $(f < c)$ is open. Let $x \in G = (f < c)^c$, that is $f(x) < c$. Let us take $\varepsilon = c - f(x) > 0$. Since f is *s.s.c*, there exists a neighborhood V of x such that

$$y \in V \Rightarrow f(y) < f(x) + \varepsilon = c,$$

which may be written as

$$y \in V \Rightarrow f(y) < c,$$

which means that $V \subseteq G^c$. We proved that G^c contains each of its elements with one of their neighborhood. Then G^c is open. This proves the direct sense.

Now suppose $(f \geq c)$ is closed for any real number c . Fix x in S . Then for any $\varepsilon > 0$, the set $G = (f < f(x) + \varepsilon)$ is open and $x \in G$. Then, there is a neighborhood of x such that $x \in V \subset G$. We conclude that : for any $x \in S$, for any $\varepsilon > 0$ there exists a neighborhood of x such that

$$y \in V \Rightarrow f(y) \leq f(x) + \varepsilon.$$

So f is **s.s.c.**. This completes the proof of Part (1).

Point (2) is proved by taking applying Point (1) to $-f$.

Point (3) is a consequence of Points (1) and (2) and classical measurability criteria for real-valued functions.

7.3. Probabilistic property of a non-countable family of disjoint events.

PROPOSITION 19. *Let $(B_\lambda)_{\lambda \in \Gamma}$ be a family of disjoint measurable sets in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then at most, a countable number of them are not null sets, or equivalently, the cardinality of the elements λ of Γ for which $\mathbb{P}(B_\lambda) > 0$, is at most countable.*

Proof. Define

$$D = \{\lambda \in \Gamma, L(B_\lambda) > 0\}.$$

and for any integer $k \geq 1$,

$$D_k = \{\lambda \in \Gamma, L(B_\lambda) > 1/k\}.$$

It is clear that we have

$$D = \cup_{k \geq 1} D_k,$$

We are going to prove that each D_k is finite. Indeed, suppose we can find $r \geq 1$ elements in D_k denoted as $\lambda_1, \lambda_2, \dots, \lambda_r$. Since the B_λ 's are disjoint, we have

$$1 \geq \mathbb{P}\left(\bigcup_{j=1}^r B_{\lambda_j}\right) = \sum_{j=1}^r \mathbb{P}(B_{\lambda_j}) \geq r/k.$$

Then

$$r \leq k.$$

This means that we cannot choose more than k points in D_k . Hence D is finite, that cardinality of D_k is less than k . Thus, D is at most countable as a countable union of finite sets

7.4. Measurability of the set discontinuity points in a metric space. Here is an amazing result, that is the sets of discontinuity points of a function defined from metric space to another metric space is measurable whatever be the function.

LEMMA 3. *Let g be a function g from the metric space (S, d) to the metric space (D, r) . Denote by $\text{discont}(g)$, the set of discontinuity points of g . We have*

$$(7.3) \quad \text{discont}(g) = \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$$

where for each couple of positive integers (s, t)

$$B_{s,t} = \{x \in S, \exists(y, z) \in S^2, d(x, y) < 1/t, d(z, x) < 1/t, r(g(y), g(z)) \geq 1/s\}.$$

is an open set.

From this lemma, we see that $\text{discont}(g)$ is measurable as countable unions and intersections of open sets. But we have to prove the lemma.

Proof of the lemma. Let us show that

$$\bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t} \subseteq \text{discont}(g).$$

Let $x \in \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$. Then there exists an integer $s \geq 1$ fixed such that for any integer $t \geq 1$, there exist y_t and z_t such that

$$d(x, y_t) < 1/t,$$

and

$$d(x, z_t) < 1/t,$$

and

$$(7.4) \quad \forall t \geq 1, \quad r(g(y_t), g(z_t)) \geq 1/s$$

Since, g is continuous at x , we get, as $t \rightarrow +\infty$,

$$r(g(y_t), g(z_t)) \leq r(g(y_t), g(x)) + r(g(x_t), g(z_t)) \rightarrow 0$$

which is in contradiction of (7.4). Then x is a discontinuity point of g .

Conversely, we have to show that

$$discont(g) \subseteq \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}.$$

Let x be a discontinuity point of g . By the negation of the definition of the continuity, we have,

$$\exists \epsilon > 0, \forall \eta > 0, \exists y \in S, \quad d(x, y) < \eta, \quad r(g(y), g(x)) \geq \epsilon.$$

Let s be an integer such that $\epsilon \geq 1/s$. Then for any $1/t$ where t is a positive integer, we have

$$\exists y \in S, \quad d(x, y) < 1/t, \quad r(g(y), g(x)) \geq 1/s.$$

Putting $z = x$, leads to

$$d(x, z) < 1/t, \quad d(x, y) < 1/t, \quad r(g(y), g(x)) \geq 1/s.$$

Then $x \in \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$.

By combining the two steps, we get the equality.

Let us prove that for each couple of positive integers (s, t) is an open set. Fix $s \geq 1, t \geq 1$. Put $a = 1/s > 0$ et $b = 1/t > 0$. Let $x \in B_{s,t}$. Then

$$\exists(y, z) \in S^2, \quad d(x, y) < b, \quad d(z, x) < b, \quad r(g(y), g(z)) \geq a$$

Set $c = \min(b - d(x, y), b - d(z, x)) > 0$ and take $x' \in B(x, c)$. Then

$$d(x', y) \leq d(x', x) + d(x, y) < c + d(x, y) \leq b$$

and next,

$$d(x', z) < d(x', x) + d(x, z) \leq c + d(x, z) \leq b$$

and

$$r(g(y), g(z)) \geq a$$

Thus, $x' \in B_{s,t}$. Hence

$$x \in B(x, c) \subseteq B_{s,t}$$

Therefore each $B_{s,t}$ contains each of its point with an open ball. Hence $B_{s,t}$ is an open set.

We finished the proof of the lemma, which proves the measurability of g .

7.5. Stone-Weierstrass Theorem. Here are two forms of Stone-Weierstrass Theorem. The second is more general and is the one we use in this text.

PROPOSITION 20. *Let (S, d) be a compact metric space and H a non-void subclass of the class $\mathcal{C}(S, \mathbb{R})$ of all real-valued continuous functions defined on S . Suppose that H satisfies the following conditions.*

(i) *H is reticulated, that is, for any couple (f, g) of elements of H , $f \wedge g$ et $f \vee g$ are in H*

(ii) *For any couple (x, y) of elements of S and for any couple (a, b) of real numbers such that $a = b$ if $x = y$, there exists a couple (h, k) of elements of H such that*

$$h(x) = a \text{ and } k(y) = b.$$

Then H is dense in $\mathcal{C}(S, \mathbb{R})$ endowed with the uniform topology, that is each continuous function from S to \mathbb{R} is the uniform limit of a sequence of elements in H .

THEOREM 7. *Let (S, d) be a compact metric space métrique and H a nonvoid subclass of the class $\mathcal{C}(S, \mathbb{C})$ of all real-valued continuous functions defined on S . Suppose that H satisfies the following conditions.*

(i) *H contains all the constant functions.*

(ii) *For all $(h, k) \in H^2$, $h + k \in H$, $h \times k \in H$, $\bar{u} \in H$.*

(iii) *H separates the points of S , i.e., for two distinct elements of S , x and y , that is $x \neq y$, there exists $h \in H$ such that*

$$h(x) \neq h(y).$$

Then H is dense in $\mathcal{C}(S, \mathbb{C})$ endowed with the uniform topology, that is each continuous function from S to \mathbb{C} is the uniform limit of a sequence of elements in H .

Remark.

If we work in \mathbb{R} , the condition on the conjugates - $\bar{u} \in H$ - becomes needless.

7.6. Divers. The min function is Lipschitz. We have for any real numbers x, y, X , and Y ,

$$(7.5) \quad |\min(x, y) - \min(X, Y)| \leq |x - X| + |y - Y|.$$

To see that, let us have a look to the four possible cases.

Case 1 : $\min(x, y) = x$ and $\min(X, Y) = X$. We have

$$|\min(x, y) - \min(X, Y)| \leq |x - X|$$

Case 2 : $\min(x, y) = x$ and $\min(X, Y) = Y$. If $x \leq Y$, we have $Y \geq X$, we have

$$0 \leq \min(X, Y) - \min(x, y) = Y - x \leq X - x$$

If $x > Y$, we have $X \geq Y$, we have

$$0 \leq \min(x, y) - \min(X, Y) = x - Y \leq y - Y$$

Case 3 : $\min(x, y) = y$ and $\min(X, Y) = Y$. We have

$$|\min(x, y) - \min(X, Y)| \leq |y - Y|$$

Case 4 : $\min(x, y) = y$ and $\min(X, Y) = X$. This case is handled as for Case 2 by permuting the roles of (x, y) and (X, Y) .

We get (7.5) by putting together the results of the four cases.

CHAPTER 3

Uniform Tightness and Asymptotic Tightness

1. Introduction

Any limit theory deals with the notion of compactness through the existence or not for sequences of subsequences converging in the sense of the defined limit. This corresponds to the Bolzano-Weierstrass for real sequences. For the weak convergence, the condition of the existence of such subsequences is called *tightness*. When dealing with weak convergence for general metric spaces, *tightness* leads to the general Prohorov theorem which establishes, under eventually other assumptions, that every *uniformly tight* sequence of measurable applications of a metric space (S, d) has at least a weakly converging subsequence.

In this chapter, we focus on weak convergence in \mathbb{R}^k . And there exists a specific handling of *weak compactness* that is very different from the treatment in the general case. In \mathbb{R}^k , the major role is played by the theorem of Helly-Bray that directly makes use of the Bolzano-Weierstrass theorem in \mathbb{R} .

Since, we deal with compact sets of \mathbb{R}^k , just remind two properties whom we are going to use. The first is that compact sets of \mathbb{R}^k are closed and bounded sets. The second is that \mathbb{R}^k is a complete and separable metric space.

Here, we will be mainly dealing with the *max*-norm defined for $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ by

$$\|x\| = \max_{1 \leq i \leq k} |x_i|.$$

The open balls $B(x, r)$ and the closed balls $B^f(x, r)$ with respect to this norm are

$$B(x, r) = \{x \in \mathbb{R}^k, \|x\| < r\} = \prod_{i=1}^k]x_i - r, x_i + r[$$

for $x = (x_1, \dots, x_k)$ and $r > 0$, and

$$B^f(x, r) = \{x \in \mathbb{R}^k, \|x\| \leq r\} = \prod_{i=1}^k [x_i - r, x_i + r]$$

for $r \geq 0$.

Before we begin, let us make some notation.

Let $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$. We define the following order relations :

$$(a \leq b) \iff (\forall (1 \leq i \leq k), a_i \leq b_i),$$

and,

$$(a < b) \iff (\forall (1 \leq i \leq k), a_i \leq b_i, \exists (1 \leq i_0 \leq k), a_{i_0} < b_{i_0})$$

and finally,

$$(a \prec b) \iff (\forall (1 \leq i \leq k), a_i < b_i,)$$

with its symmetrical counterpart,

$$(a \succ b) \iff (\forall (1 \leq i \leq k), a_i > b_i,)$$

Also, let us define the following classes of compact sets.

For $A = (A_1, \dots, A_k) \prec V = (B_1, \dots, B_k)$, denote

$$K_{A,B} = \prod_{i=1}^k [A_i, B_i].$$

For $A = (A_1, \dots, A_k) \succ 0$, set

$$K_A = \prod_{i=1}^k [-A_i, A_i].$$

For $M \in \mathbb{R}$, $M > 0$, put

$$K_{c,M} = [-M, M]^k.$$

The sets $K_{A,B}$, K_A and $K_{c,M}$, are compacts and will be used to characterize the tightness of sequences. The next proposition paves the way for the statements of different and equivalent conditions for tightness.

PROPOSITION 21. *Let $\{\mathbb{P}_n, n \geq 1\}$ be a sequence of probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The following propositions are equivalent.*

(1a) *For any $\varepsilon > 0$, there exists a compact set K in \mathbb{R}^k such that*

$$\inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2a) *For any $\varepsilon > 0$, there exists a real number $M > 0$ such that*

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

(3a) *For any $\varepsilon > 0$, there exists a vector $A = (A_1, \dots, A_k) \succ 0$ of \mathbb{R}^k such that*

$$\inf_{n \geq 1} \mathbb{P}_n(K_A) \geq 1 - \varepsilon.$$

(4a) *For any $\varepsilon > 0$, there exist two vectors $A = (A_1, \dots, A_k) \prec B = (B_1, \dots, B_k)$ of \mathbb{R}^k such that*

$$\inf_{n \geq 1} \mathbb{P}_n(K_{A,B}) \geq 1 - \varepsilon.$$

(1b) *For any $\varepsilon > 0$, there exists a compact set K of \mathbb{R}^k such that*

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2b) *For any $\varepsilon > 0$, there exists a real number $M > 0$ such that*

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

(3b) *Pour tout $\varepsilon > 0$, il existe un vecteur $A = (A_1, \dots, A_k) \succ 0$, de \mathbb{R}^k tels que*

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_A) \geq 1 - \varepsilon.$$

(4b) *For any $\varepsilon > 0$, there exists two vectors $A = (A_1, \dots, A_k) \prec B = (B_1, \dots, B_k)$ of \mathbb{R}^k such that*

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_{A,B}) \geq 1 - \varepsilon.$$

Proof. We have two groups of formulas : (1a) – (4a) and (1b) – (4b). In fact, we are going to prove that the different points of each group are equivalent and next, that the two first points of the two groups are.

Equivalence between the points of the first group (1a)-(4b):

Let $\varepsilon > 0$ be fixed. Let us show :

(1a) \implies (2a). Let K be a compact set such that $\sup_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon$. Since K is compact, it is bounded. Then, it is included in a set of the form $\{x, \|x\| \leq M\} = K_{c,M}$ and then

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq \inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2a) \implies (3a). This is obvious since $K_{c,M}$ is equal to K_A with $A = (M, M, \dots, M)$.

(3a) \implies (4a). This is also obvious since a set of the form K_A , for $A = (A_1, \dots, A_k) \succ 0$, is exactly $K_{-A,A}$.

(4a) \implies (1a). This is also obvious since $K_{A,B}$ is a compact set of \mathbb{R}^k .

Equivalence between the points of the group (1b)-(4b). The proof is exactly the same as for the first group.

Equivalence between the two groups. It will be enough to prove that : (1a) \iff (1b).

If (1a) holds, then for any $\varepsilon > 0$, there exists a compact set K of \mathbb{R}^k such that, for any $n \geq 1$,

$$\mathbb{P}_n(K) \geq 1 - \varepsilon.$$

Therefore, we have

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

This leads to (1b).

If (1b) holds, Then for any $\varepsilon > 0$, there exists a compact set K such that

$$\left\{ \sup_{n \geq 1} \inf_{p \geq n} \mathbb{P}_p(K) \right\} \geq 1 - \varepsilon/2.$$

Then, there exists $N \geq 1$, such that

$$\inf_{p \geq N+1} \mathbb{P}_p(K) \geq 1 - \varepsilon,$$

that is for any $n > N$,

$$\mathbb{P}_n(K) \geq 1 - \varepsilon.$$

Since K is a compact set, it is in a set of the form K_{c,M_∞} , where $M_\infty > 0$, and then, for any $n > N$,

$$\mathbb{P}_n(K_{c,M_\infty}) \geq 1 - \varepsilon.$$

Now, for each fixe j , $1 \leq j \leq N$, the set $(\|x\| \leq M) = K_{c,M}$ increases with A to \mathbb{R}^k and then, $\mathbb{P}(\|X_j\| \leq M) \uparrow 1$. Thus, for any $1 \leq j \leq N$, there exists a real number $M_j > 0$,

$$\mathbb{P}_j(K_{c,M_j}) \geq 1 - \varepsilon.$$

By passing, we just demonstrated that each probability measure $\mathbb{P}^{(0)}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is **tight**, that is for any $\varepsilon > 0$, there exists a compact set $K^{(0)} = K_{c,M^{(0)}}$ in \mathbb{R}^k such that

$$(1.1) \quad \mathbb{P}^0(K^{(0)}) \geq 1 - \varepsilon.$$

Coming back to our proof, we may take

$$M = \max(M_1, \dots, M_N, M_\infty),$$

and see that the sets K_{c,M_j} , $1 \leq j \leq M$ and K_{c,M_∞} are all in $K_{c,M}$ and then for $n \geq 1$,

$$\mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon$$

and donc

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon,$$

which is (1a), since $K_{c,M}$ is a compact set.

In a new step, we provide a link between the formulas and distribution functions. For a reminder, recall that the probability distribution function associated with a probability measure \mathbb{P} is defined by

$$F_{\mathbb{P}}(x) = \mathbb{P}([-\infty, x]) = \mathbb{P}\left(\prod_{i=1}^k [-\infty, x_i]\right), x = (x_1, \dots, x_k) \in \mathbb{R}^k.$$

This probability distribution function, in turn, determines the probability measure \mathbb{P} as the Lebesgues-Stieljes probability measure defined by : for any $(a, b) \in \mathbb{R}^k \times \mathbb{R}^k$, $a \leq b$,

$$\mathbb{P}([a, b]) = \Delta_{a,b}F = \sum_{\varepsilon \in \{0,1\}^k} (-1)^{s(\varepsilon)} F(b + \varepsilon * (a - b)) \geq 0,$$

where for $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{0,1\}^k$, $s(\varepsilon) = \varepsilon_1 + \dots + \varepsilon_k$, pour $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, $y = (y_1, \dots, y_k)$, $x * y = (x_1 y_1, \dots, x_k y_k)$.

We are going to use the Lebesgues-Stieljes probability measures to deal with uniform tightness. The reader is directed to [9] or specially to the

Chapter 1 of [8].

For now, we need this notation. Denote for $M > 0$.

$$L_M = \{x, \exists(1 \leq i \leq k), x_i \leq -c\}$$

We have

PROPOSITION 22. *Let $\{\mathbb{P}_n, n \geq 1\}$ be a sequence of probability measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and consider the sequence of their probability distribution functions $\{F_n \geq 1\}$ with $F_{\mathbb{P}_n} = F_n$ for $n \geq 1$. Then the three following points are equivalent.*

(1c) *For any $\varepsilon > 0$, there exist a vector $0 < C \in \mathbb{R}^k$ and a real number $c > 0$ such that*

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon$$

and

$$\inf_{n \geq 0} \mathbb{P}_n(L_c) \leq \varepsilon.$$

(2c) *For any $\varepsilon > 0$, there exists $0 < c$ such that for $c^{(k)} = (c, \dots, c)$, there exists $M > 0$ such that*

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon$$

and

$$\sup_{n \geq 0} \mathbb{P}_n(L_M) \leq \varepsilon.$$

(3c) *For any $\varepsilon > 0$, there exists $M > 0$ such that*

$$\inf_{n \geq 1} P_n(K_{c,M}) \geq 1 - \varepsilon.$$

Since Point (3c) is also Point (2c) of Proposition 21, then Points (3a) and (3b) are equivalent to all points of that proposition.

Proof. Let us proceed to the proofs of the different equivalence assertions.

(a) (1c) \implies (2c). For any $\varepsilon > 0$, there exists $0 < C \in \mathbb{R}^k$ such that

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon.$$

Set $c = \max\{C_i, 1 \leq i \leq k\}$. We have $] - \infty, C] \subset] - \infty, c^{(k)}]$ and $F_n(c^{(k)}) \geq F_n(C)$,

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon.$$

This finishes the proof of this step (a), since the second formula implies all the others.

(b)(2c) \implies (3c). From (2c), we find a vector $d^{(k)} = (d, \dots, d)$, with $d > 0$, such that

$$\inf_{n \geq 1} F_n(d^{(k)}) \geq 1 - \varepsilon/2$$

and real number $e > 0$ such that

$$\sup_{n \geq 1} \mathbb{P}_n(L_e) \leq \varepsilon/2.$$

By putting $M = \max(d, e)$, we get

$$\inf_{n \geq 1} F_n(M^{(k)}) \geq 1 - \varepsilon/2.$$

and next

$$\sup_{n \geq 1} \mathbb{P}_n(L_M) \leq \varepsilon/2.$$

Now, let us split \mathbb{R}^k as $\mathbb{R}^k = L_M + L_M^c$, with

$$L_M^c = \{x, \forall (1 \leq i \leq k), x_i \geq -M\},$$

which itself may be decomposed as

$$\begin{aligned} L_M^c &= \{x, \forall (1 \leq i \leq k), -M \leq x_i \leq M\} \\ &+ \{x, \forall (1 \leq i \leq k), x_i \geq -M \text{ et } \exists (1 \leq i \leq k), x_i > M\} \\ &= K_{c,M} + B, \end{aligned}$$

where, obviously,

$$B \subset]-\infty, M^{(k)}]^c.$$

Therefore, we infer from $\mathbb{R}^k = L_M + K_{c,M} + B$ that

$$(1.2) \quad K_{c,M}^c = L_M + B..$$

Thus, for any $n \geq 1$,

$$\mathbb{P}_n(K_{c,M}) = \mathbb{P}_n(L_M) + \mathbb{P}_n(B) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

since $B \subset]-\infty, M^{(k)}]^c$. Hence for any $n \geq 1$,

$$\begin{aligned} \mathbb{P}_n(B) &\leq \mathbb{P}_n(]-\infty, M^{(k)}]^c) \\ &\leq 1 - \mathbb{P}_n(]-\infty, M^{(k)}]) \\ &\leq 1 - F_n(M^{(k)}) \leq \varepsilon/2. \end{aligned}$$

This ends the proof of this step (b).

(c)(3c) \implies (1c). Suppose that (3c) holds : for any $\varepsilon > 0$, there exists $M > 0$ such that

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

Then we have

$$\inf_{n \geq 1} F_n(M^{(k)}) = \inf_{n \geq 1} \mathbb{P}_n([-\infty, M^{(k)}]) \geq \inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

Next, because of (1.2), we get

$$\mathbb{P}_n(L_M) \leq \mathbb{P}_n(K_{c,M}^c) \leq \varepsilon.$$

Then (1c) holds. Proposition is entirely proved.

We move to the study of the tightness concept.

2. Tightness

2.1. Simple Tightness. In our particular case, each probability measure on $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ is tight in the following meaning.

DEFINITION 2. *A probability measure \mathbb{P} on a metric space (S, d) is tight if and only if for any $\varepsilon > 0$, there exists a compact set in S such that*

$$\mathbb{P}(K) \geq 1 - \varepsilon.$$

By Formula (1.1) below, we have

PROPOSITION 23. *A probability measure \mathbb{P} on $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ is tight.*

This result is extensible to complete and separable metric spaces, more generally to totally bounded metric spaces.

2.2. Asymptotic tightness. Uniform tightness.

DEFINITION 3. (a) *A sequence of probability measures $\{\mathbb{P}_n, n \geq 1\}$ on $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ is asymptotically tight or is uniformly tight if and only if for $\varepsilon > 0$, there exists a compact set K in \mathbb{R}^k such that*

$$(2.1) \quad \inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon$$

or, equivalently,

$$(2.2) \quad \liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(b) A sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, is asymptotically tight or is uniformly tight if and only if the sequence of the probability laws $\{\mathbb{P}_{X_n}, n \geq 1\}$ is asymptotically tight or is uniformly tight, that is for any ε , there exists a K in \mathbb{R}^k such that

$$\inf_{n \geq 1} \mathbb{P}_n(X_n \in K) \geq 1 - \varepsilon$$

or equivalently,

$$\lim_{n \rightarrow \infty} \inf \mathbb{P}_n(X_n \in K) \geq 1 - \varepsilon.$$

(c) A sequence of probability distribution functions $\{F_n, n \geq 1\}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is asymptotically tight or is uniformly tight if and only if the sequence of their Lebesgues-Stieljes probability measures $\{\mathbb{P}_n, n \geq 1\}$ is asymptotically tight or is uniformly tight, or equivalently for any $\varepsilon > 0$, there exist $0 < C \in \mathbb{R}^k$ and a real number $c > 0$ such that

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon$$

and

$$\sup_{n \geq 0} \mathbb{P}_n(L_c) \leq \varepsilon.$$

if and only if, there exists for any $\varepsilon > 0$, a real number $c > 0$ such that we have for $c^{(k)} = (c, \dots, c)$

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon.$$

In \mathbb{R}^k , uniform tightness (2.1) is equivalent to asymptotic tightness because of Proposition 21. Thus from now, we speak about only about tightness of sequences of probability measures, or of random vectors, or of probability distribution functions.

Before, we come to the Helly-Bray theorem, we are going to give three important properties of tightness.

2.3. Tightness and continuous mapping. The tightness is preserved by continuous mapping in the following sense.

PROPOSITION 24. *Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, be a tight sequence of random vectors and let $g : \mathbb{R}^k \mapsto \mathbb{R}^m$, $m \geq 1$, be a continuous mapping. Then the sequence $\{g(X_n), n \geq 1\}$ is tight.*

Proof. Let $\{X_n, n \geq 1\}$ be tight and $g : \mathbb{R}^k \mapsto \mathbb{R}^m$ continuous. For any $\varepsilon > 0$, there exists a compact set K in \mathbb{R}^k such that

$$(2.3) \quad \inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

But $(X_n \in K) \subset (g(X_n) \in g(K))$ where

$$K_0 = g(K) = \{g(x), x \in K\}$$

is the direct image of K by g , and is a compact set. Indeed, let $\{g(x_n), x_n \in K, n \geq 1\}$ be a sequence in K_0 . Since K is a compact set, the sequence $(x_n)_{n \geq 1}$, which is in K , has a subsequence $x_{n(k)} \rightarrow x \in K$ converging, as $k \rightarrow +\infty$, to a point x which is in K since K is closed. Since g is continuous, then $g(x_{n(k)})$ converges to $g(x) \in K_0$ as $k \rightarrow +\infty$. It follows that K_0 is a compact set in \mathbb{R}^m and

$$\inf_{n \geq 1} \mathbb{P}(g(X_n) \in K_0) \geq \inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

This ends the proof.

2.4. Characterization of the tightness by that of the components. In the particular case of \mathbb{R}^k , we have

PROPOSITION 25. *A sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, is tight if and only if each sequence of components, $\{X_n^{(i)}, n \geq 1\}$, $1 \leq i \leq k$, is tight.*

Proof. Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, be a sequence of random vectors.

Suppose that this sequence is tight. By Proposition 24, each sequence of components $\{X_n^{(i)}, n \geq 1\} = \{\pi_i(X_n), n \geq 1\}$, $1 \leq i \leq k$, is tight, as continuous transformations of a tight sequence, that is as the i -th projection π_i of a tight sequence.

Suppose that for each $1 \leq i \leq k$, $\{X_n^{(i)}, n \geq 1\}$ is tight. Then for $1 \leq i \leq k$, for any $\varepsilon > 0$, there exists a real number $A_i > 0$ such

$$\inf_{n \geq 1} \mathbb{P}(X_n^{(i)} \in [-A_i, A_i]) \geq 1 - \varepsilon/k.$$

By setting $A = (A_1, \dots, A_k)$, we have for $A > 0$

$$\bigcap_{i=1}^k (X_n^{(i)} \in [-A_i, A_i]) = \left(X_n \in \prod_{i=1}^k [-A_i, A_i] \right),$$

It comes that for any $n \geq 1$,

$$\begin{aligned} \mathbb{P} \left(X_n \notin \prod_{i=1}^k [-A_i, A_i] \right) &= \mathbb{P} \left(\bigcup_{i=1}^k (X_n^{(i)} \notin [-A_i, A_i]) \right) \\ &\leq \sum_{i=1}^k P(X_n^{(i)} \notin [-A_i, A_i]) \leq \varepsilon, \end{aligned}$$

and then for any $n \geq 1$,

$$\mathbb{P}(X_n \in K_A) \geq 1 - \varepsilon.$$

Hence, the sequence $\{X_n, n \geq 1\}$ is tight. The proof is complete.

2.5. Tightness of a weakly convergence sequence.

PROPOSITION 26. *Any sequence $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, be a sequence of random vectors that weakly converges is tight.*

Proof. Suppose that X_n weakly converges to the probability \mathbb{P} . This probability is tight. So for any $\varepsilon > 0$, there exists a compact $K_A = [-A, A]$ of K such that

$$\mathbb{P}(K) \geq 1 - \varepsilon.$$

Let $0 < \delta < 1$ and set $A + \delta = (A_1 + \delta, \dots, A_k + \delta)$. We have for any $0 < \delta < 1$,

$$\overset{\circ}{K}_{A+\delta} = \prod_{i=1}^k]-A_i - \delta, A_i + \delta[.$$

Since $X_n \rightsquigarrow X$ and $\overset{\circ}{K}_{A+\delta}$ is open, we use Point (ii) of Portmanteau Theorem 2 to show that

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \liminf_{n \rightarrow \infty} P(X_n \in \overset{\circ}{K}_{A+\delta}) \geq \mathbb{P}(\overset{\circ}{K}_{A+\delta}),$$

for any $0 < \delta < 1$ and next

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \mathbb{P}(\overset{\circ}{K}_{A+\delta}),$$

for any $0 < \delta < 1$. By letting $\delta \downarrow 0$, we have $\overset{\circ}{K}_{A+\delta} \downarrow \overline{K} = K$ since K is a closed set. Applying this in the last formula gives

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \mathbb{P}(K) \geq 1 - \varepsilon.$$

It comes that the sequence $\{X_n, n \geq 1\}$ is tight.

Remark. Actually, we may see that the sequence has inherited the tightness of the weak limit. This result still holds for complete and

separable spaces where any probability measure is tight.

In the new section, we are going to deal with the fundamental theorem of tightness.

3. Compacity Theorem for weak convergence in \mathbb{R}^k

This theorem is a kind of inverse of Proposition 26, concerning the convergence of subsequence.

THEOREM 8. (*Prohorov - Helly-Bray*) *Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, be a tight sequence of random vectors. Then it contains a weakly converging subsequence.*

This theorem may be directly proved, as done in [2] and van der Vaart and Wellner [11]. The proof in Billingsley is very lengthy. That of van der Vaart and Wellner is very much simpler and more general. But in this context, we are going to use the Helly-Bray approach as in van der Vaart [12] and Loève [10]. Here, we give a more detailed proof.

Here, the proof of Theorem 8 is based on the following Helly-Bray Theorem in which the hard work is done.

THEOREM 9. (*Helly-Bray*) *Any sequence $\{F_n, n \geq 1\}$ of probability distribution function on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ has a subsequence $\{F_{n(k)}, k \geq 1\}$ weakly converging to a distribution function F , which is not necessarily a probability distribution function.*

Proof. Let $\{F_n, n \geq 1\}$ be a sequence of probability distribution functions $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Let \mathbb{Q}^k be the set of all elements of \mathbb{R}^k with rational components. \mathbb{Q}^k is everywhere dense in \mathbb{R}^k . Let us enumerate \mathbb{Q}^k as $\mathbb{Q}^k = \{q_1, q_2, \dots\}$ and proceed by steps.

Step 1. We are going to find a subsequence $(F_{n(j)})_{j \geq 1}$ of $(F_n)_{n \geq 1}$ pointwise converging to some function G on \mathbb{Q}^k by using the diagonal sequence method. We have that $(F_n(q_1))_{n \geq 1} \subset [0, 1]$. Using Bolzano-Weierstrass property on \mathbb{R} , we get a subsequence $(F_{1,n}(q_1))_{n \geq 1}$ of $\{F_n - q_1, n \geq 1\}$ converging to $G(q_1)$.

Next, we apply the subsequence $(F_{1,n})_{n \geq 1}$ to q_2 in that way : $(F_{1,n}(q_2))_{n \geq 1} \subset [0, 1]$. We find a subsequence $(F_{2,n}(q_2))_{n \geq 1}$ of $(F_{1,n}(q_2))_{n \geq 1}$ that converges to a real number $G(q_2)$. We proceed so forth and get subsequences $(F_{j,n})_{n \geq 1}$, $j = 1, 2, \dots$ satisfying :

(a) For each $j \geq 1$, $(F_{j+1,n})_{n \geq 1}$ is a subsequence of any of the subsequences $(F_{i,n})_{n \geq 1}$ $1 \leq i \leq j$.

(b) For any $j \geq 1$, for any $1 \leq j \leq i$, $F_{j,n}(q_i) \rightarrow G(q_i)$.

Next, we take the diagonal sequence $(F_{j,j})_{j \geq 1}$. We may use a simple graph, as below, to see this : for any fixed $i \geq 1$, the sequence $\{F_{j,j}, j \geq i\}$ is a subsequence of $(F_{i,n})_{n \geq i}$ and then

$$F_{j,j}(q_i) \rightarrow G(q_i).$$

To read this graph, one has to notice that the sequence in one line is a subsequence of those in the preceding lines. From this, it becomes clear that $\mathbf{F}_{j,j}$ an element of all the lines from 1 to j .

$\mathbf{F}_{1,1}$	$F_{1,2}$	$F_{1,3}$	$F_{1,4}$	$F_{1,5}$	$F_{1,6}$	$F_{1,7}$	$F_{1,8}$	$F_{1,9}$	$F_{1,10}$	\dots
$\mathbf{F}_{2,2}$	$F_{2,3}$	$F_{2,4}$	$F_{2,5}$	$F_{2,6}$	$F_{2,7}$	$F_{2,8}$	$F_{2,9}$	$F_{2,10}$	$F_{2,11}$	\dots
	$\mathbf{F}_{3,3}$	$F_{3,4}$	$F_{3,5}$	$F_{3,6}$	$F_{3,7}$	$F_{3,8}$	$F_{3,9}$	$F_{3,10}$	$F_{3,11}$	\dots
		$\mathbf{F}_{4,4}$	$F_{4,5}$	$F_{4,6}$	$F_{4,7}$	$F_{4,8}$	$F_{4,9}$	$F_{4,10}$	$F_{4,11}$	\dots
			\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
				$\mathbf{F}_{j,j}$	$F_{j,j+1}$	$F_{j,j+2}$	$F_{j,j+3}$	$F_{j,j+4}$	$F_{j,j+5}$	\dots

We conclude that the diagonal subsequence $(F_{j,j})_{j \geq 1}$, written as $(F_{n(j)})_{j \geq 1}$, satisfies

$$\forall q \in \mathbb{Q}^k, F_{n(j)}(q) \rightarrow G(q) \text{ quand } j \rightarrow +\infty.$$

Step 2. Properties of G on \mathbb{Q}^k .

(2.1) For any $(a, b) \in \mathbb{Q}^k \times \mathbb{Q}^k$, as $j \rightarrow +\infty$,

$$\begin{aligned} 0 \leq \Delta_{a,b} F_{n(j)} &= \sum_{\epsilon \in \{0,1\}^k} (-1)^{s(\epsilon)} F_{n(j)}(b + \epsilon * (a - b)) \\ &\rightarrow \Delta_{a,b} G = \sum_{\epsilon \in \{0,1\}^k} (-1)^{s(\epsilon)} G(b + \epsilon * (a - b)) \geq 0, \end{aligned}$$

Since all the points $b + \epsilon * (a - b)$ are in \mathbb{Q}^k . It comes that G has nonnegative rectangles on \mathbb{Q}^k .

G is nondecreasing on \mathbb{Q}^k as inherited from the nondecreasingness of the $F_{n(j)}$, $j \geq 1$, on \mathbb{Q}^k .

step 3. Define F on $\mathbb{J}^k = (\mathbb{R} \setminus \mathbb{Q})^k$ by

$$F(x) = \inf\{G(q), q \in \mathbb{Q}^k, x \prec q\} \in [0, 1].$$

for $x \in \mathbb{J}^k$. It is obvious that F is well-defined on \mathbb{J}^k . It is also sure that F is non-decreasing.

(a) Let us show that F is right-continuous. Let $x \in \mathbb{J}^k$ and let $\varepsilon > 0$. By definition of the finite infimum, there exists $q \in \mathbb{Q}^k$ such that $x \prec q$ and $G(q) < F(x) + \varepsilon$. For any $y \in \mathbb{J}^k$, $x \prec y < q$, we have $F(y) \leq G(q)$ and $\varepsilon > G(q) - F(x) \geq F(y) - F(x)$. Then

$$(3.1) \quad \forall \varepsilon > 0, \exists q > x, x < y < q \implies 0 \leq F(y) - F(x) < \varepsilon.$$

Then F is right-continuous à droite.

(c) Let us show that $F_{n(j)}(x) \rightarrow F(x)$ for continuity points of $x \in \mathbb{J}^k$ of G .

Let x be a continuity point of F on $x \in \mathbb{J}^k$. For any $\varepsilon > 0$, we may find (y', y'') in \mathbb{J}^k such that $y' < x < y''$ and $F(y'') - F(y') < \varepsilon/2$. Soit $(q', q'') \in \mathbb{Q}^k$ such that $y' < q' < x < q'' < y''$. Then $G(q'') - F(q') \leq F(y'') - F(y') \leq \varepsilon$. Next

$$\begin{aligned} F(y') \leq G(q') = \lim F_{n(j)}(q') &\leq \liminf F_{n(j)}(x) \leq \limsup F_{n(j)}(x) \\ &\leq \limsup F_{n(j)}(q'') = \lim G(q'') \leq F(y''). \end{aligned}$$

Then $\liminf_{j \rightarrow +\infty} F_{n(j)}(x)$, $\limsup_{j \rightarrow +\infty} F_{n(j)}(x)$ and $F(x)$ are in the interval $[F(y'), F(y'')]$ with length at most equal to ε . This implies that

$$(3.2) \quad \max(|F(x) - \liminf F_{n(j)}(x)|, |F(x) - \limsup F_{n(j)}(x)|) \leq \varepsilon,$$

for any $\varepsilon > 0$. Therefore, we arrive at

$$F_{n(j)}(x) \rightarrow F(x) \text{ as } j \rightarrow +\infty.$$

(d) F has nonnegative rectangles. For any $(a, b) \in \mathbb{J}^k \times \mathbb{J}^k$, let $q' \downarrow a$ and $q'' \downarrow b$ with $q' \succ a$ and $q'' \succ b$ and $(q', q'') \in \mathbb{Q}^k \times \mathbb{Q}^k$. By

monotone limit, and by the definition of G ,

$$(3.3) \quad 0 \leq \Delta_{q', q''} G \rightarrow \Delta_{a, b} F \geq 0.$$

Partial conclusion. G is a distribution function on \mathbb{J}^k and $F_{n(j)}(x) \rightarrow F(x)$ for continuity points $x \in \text{of } G$.

Step 4. Now, we may extend F on $\mathbb{J}_c^k = \mathbb{R}^k \setminus \mathbb{J}^k$ by

$$F(x) = \inf\{F(y), y \in \mathbb{J}^k, x \prec y\} \in [0, 1], \quad x \in \mathbb{J}_c^k.$$

First, with a very little effort, we see that F is non-decreasing on \mathbb{R}^k . Next, we have to prove that F is right-continuous at any point $x \in \mathbb{R}^k$. Let us fix $\varepsilon > 0$. If $x \in \mathbb{J}^k$, by right-continuity of F on \mathbb{J}^k , we can find $y_1 \succ x$, for any $z \in \mathbb{J}^k$ and $x \leq z \prec y_1$, we have : $G(x) \leq G(z) \leq G(y_1) < G(x) + \varepsilon$. This is also true for $z \in \mathbb{J}_c^k$ since, by construction, $G(z) \leq G(y_1)$.

If $x \in \mathbb{J}_c^k$, by definition of $F(x)$ we can find $y_1 \succ x$, such that for any $z \in \mathbb{J}^k$ and $x \leq z \prec y_1$, we have $G(x) \leq G(z) \leq G(y_1) < G(x) + \varepsilon$. We conclude as in the first case by using in addition the increasingness of F . We conclude that for any \mathbb{R}^k , for any $\varepsilon > 0$, we can find $y \succ x$ in \mathbb{J}^k such that

$$(x \leq z \prec y) 0 < G(z) - G(x) < \varepsilon G(y) - G(x) < \varepsilon.$$

Thus, F is right-continuous.

Finally, to show that F has positive rectangles, we use the right continuity of F and the fact that the property holds on \mathbb{J}^k .

Finally, we have to prove that $F_{n(j)}(x) \rightarrow F(x)$ for any continuity points of F . We may repeat the same technique that led to (3.2). We are going to give only the beginning.

Let x be a continuity of F on $x \in \mathbb{R}^k$. For any $\varepsilon > 0$, we may find (y', y'') in \mathbb{R}^k such that $y' \prec x \prec y''$ and $F(y'') - F(y') < \varepsilon/2$.

From there, we may find $(z', z'') \in \mathbb{J}^k$ such that $y' < z' < x < z'' < y''$ and such that z' and z'' are continuity points of F on \mathbb{J}^k . If we are able to do that, we may recondut the same lines that led to (3.2), by replacing (y', y'') by (z', z'') .

Now, we may find points of the form $z_\varepsilon = y' + (\varepsilon)\delta$ in $]y', x[\cap \mathbb{J}^k$, where $\delta = (1, \dots, 1)$ and $0 < \varepsilon < \varepsilon_0$. The boundaries of the intervals $] -\infty, z_\varepsilon]$

are disjoint. So, for $m_{F,\mathbb{J}}$ being the Lebesgues-Stieljes measure associated with F on \mathbb{J}^k , we may have $m_{F,\mathbb{J}}(\partial] - \infty, z_\varepsilon]) > 0$ only for - at most - a countable number of z_ε . Then we may easily pick a value of ε such that $m_{F,\mathbb{J}}(\partial] - \infty, z_\varepsilon]) > 0$, that is $z' = z_\varepsilon$ is a continuity point of F on \mathbb{J}^k . We find z'' in the same manner.

This completely finishes the proof.

Remark We wanted to give a complete proof with all the necessary details. Our step 4 is needless if it is possible to prove that G is right-continuous on \mathbb{Q}^k . If this is the case, one should stop at Step 3 and take $F = G$.

Now let us move to the proof of Theorem 8.

Proof of Theorem 8 of Prohorov. Suppose that the sequence of probability distribution functions $\{F_n, n \geq 1\}$ is tight, that is the sequence of their Lebesgues-Stieljes measures $\{\mathbb{P}_n(\cdot) = \Delta_{a,b}F_n, n \geq 1\}$ is tight. By Proposition 22, for any $\varepsilon > 0$, we may find a vector $C \succ 0$, $C \in \mathbb{R}^k$ such that $n \geq 1$,

$$F_n(C) \geq 1 - \varepsilon.$$

By Theorem 9, there exists a subsequence $(F_{n(j)})_{j \geq 1}$ of $(F_n)_{n \geq 1}$ that weakly converges to a distribution function F associated to a mesure L defined by $L(\cdot) = \Delta_{a,b}F$ and bounded by the unity.

Consider the family $\{C_h = C + h^{(k)}, h > 0\}$. These points are such that their boudaries are $\partial] - \infty, F = C_h]$ are disjoint. So, we may choose a sequence C_{h_p} such that $L(\partial] - \infty, C_{h_p}) = 0$ for any $p \geq 1$ and $C_{h_p} \uparrow (+\infty)^{(k)}$ as $p \uparrow +\infty$. These points are continuity ones of F and are greater than C . Then for any fixed $p \geq 1$,

$$F_{n(j)}(C_{h_p}) \geq 1 - \varepsilon.$$

By letting $j \rightarrow \infty$, we get

$$F(C_{h_p}) = L(\cdot) = \Delta_{a,b}F \geq 1 - \varepsilon.$$

Next by letting $p \uparrow +\infty$ and next $\varepsilon \downarrow 0$, we get

$$(3.4) \quad F((+\infty)^{(k)}) = 1.$$

On another hand, for any $\varepsilon > 0$, there exists $M > 0$, such that

$$\sup \mathbb{P}_n(L_M) \leq \varepsilon.$$

We have to prove that

$$(3.5) \quad \lim_{\exists(1 \leq i \leq k), x_i \rightarrow -\infty} F(x) = 0,$$

which is equivalent to saying that for any $\varepsilon > 0$, there exists $M > 0$ such that

$$\exists(1 \leq i \leq k), x_i < -M \implies F(x) \leq \varepsilon.$$

But

$$\exists(1 \leq i \leq k), (x_i < -M) \implies (]-\infty, x] \subset L_M),$$

and then for any $n \geq 1$,

$$\exists(1 \leq i \leq k), (x_i < -M) \implies (F_n(x) \leq \varepsilon).$$

Now, let x be fixed such that : $\exists(1 \leq i \leq k), x_i < -M$. Let $x(h) = x + h^{(k)}$, with $0 < h < -(M + x_i)$. By the now classical method we used just above, we may find a sequence $x(h_p)$, $p \geq 1$, of continuity points of F with $h_p \downarrow 0$. Then for any fixed $p \geq 1$, for any $j \geq 1$,

$$F_{n(j)}(x(h_p)) \leq \varepsilon.$$

By letting $j \rightarrow \infty$, we get

$$F(x(h_p)) \leq \varepsilon.$$

Now, by right continuity, we get, as $p \uparrow +\infty$,

$$F(x) \leq \varepsilon.$$

We conclude that for any $\varepsilon > 0$,

$$\exists(1 \leq i \leq k), x_i < -M \implies F(x) \leq \varepsilon.$$

And this proves (3.5). This combined to (3.4) shows that F is a probability distribution function on \mathbb{R}^k .

4. Applications

4.1. Continuity Theorem of Lévy.

THEOREM 10. *Let ψ_n , $n \geq 1$, be a sequence of characteristic functions on \mathbb{R} that converges pointwisely to a function ψ which is continuous at zero. Then ψ is a characteristic function.*

Proof. We necessarily have $\psi(0) = 0$ since $\psi_n(0) = 0$ for all $n \geq 1$. We may suppose that the ψ_n are the characteristic functions of random variables X_n , that is for any $n \geq 1$,

$$\psi_n(t) = E(e^{itX_n}), t \in \mathbb{R}.$$

By Fact 1 in Section 2 in Chapter 6, we have that for $|\sin a| \leq a$ for $|a| \geq 2$. Then

$$1_{(|\delta x| > 2)} \leq 2(1 - \frac{\sin \delta x}{\delta x})$$

and by the right equality easily proved,

$$1_{(|\delta x| > 2)} \leq 2(1 - \frac{\sin \delta x}{\delta x}) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos tx) dt.$$

Let us apply this formula to X_n to get

$$1_{(|X_n| > 2/\delta)} \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos tX_n) dt$$

and by taking expectations and by applying Fubini Theorem for integrable functions,

$$P(|X_n| > \frac{2}{\delta}) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} R_e(1 - Ee^{itX_n}) dt.$$

By applying the Monotone Convergence Theorem, to $R_e(1 - Ee^{itX_n}) \rightarrow R_e(1 - \psi(t))$, we arrive at

$$(4.1) \quad \liminf_{n \rightarrow \infty} P(|X_n| > \frac{2}{\delta}) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} R_e(1 - \psi(t)) dt.$$

The *real part* function $R_e(\cdot)$ is continuous and by the assumptions, $R_e(1 - \psi(t)) \rightarrow 0$ as $\delta \rightarrow 0$. Therefore, we may let $\delta \rightarrow 0$ in (4.1) to get

$$\liminf_{n \rightarrow \infty} P(|X_n| > \frac{2}{\delta}) = 0.$$

This implies that the sequence is tight. Then there exists a subsequence X_{n_k} weakly converging to X . By Theorem , we have

$$\psi_{n_k}(t) = E(\exp(itX_{n_k})) \rightarrow E(\exp(itX)) = \psi_0(t).$$

By the unicity of limits in \mathbb{R}

$$\psi = \psi_0$$

We conclude that ψ is a characteristic function.

By applying this, we will have another proof of the characterization of weak convergence by characteristic functions.

4.2. Another proof of the characterization of weak convergence by characteristic functions.

THEOREM 11. *A sequence X_n of random vectors with values in \mathbb{R}^k weakly converges to the random vector $X \in \mathbb{R}^k$ if and only if $u \in \mathbb{R}^k$, $\mathbb{E}(\exp(i \langle u, X_n \rangle)) \rightarrow \mathbb{E}(\exp(i \langle u, X \rangle))$ as $n \rightarrow +\infty$.*

Proof. The direct implication comes from the application of the Dominated Convergence Theorem. Let us prove the indirect implication. Suppose that for any $u \in \mathbb{R}^k$

$$\psi_n(u) = E(\exp(i \langle u, X_n \rangle)) \rightarrow E(\exp(i \langle u, X \rangle)) = \psi_n(u).$$

For any fixed i , $1 \leq i \leq k$, the sequence of the i -th components, $X_n^{(i)}$, satisfies, for any $t \in \mathbb{R}$,

$$\psi_{n^{(i)}}(t) = \underbrace{\psi_n(0, \dots, t, \dots, 0)}_{i\text{-th place}} = E(\exp(itX_n^{(i)})) \rightarrow \underbrace{\psi(0, \dots, t, \dots, 0)}_{i\text{-th place}} = \psi^{(i)}(t).$$

The function ψ_i is continuous since it is the characteristic function of X_i . Then each $X_n^{(i)}$ is tight. By Proposition 25, the sequence X_n is tight.

We may conclude in two steps.

Step 1 : Each subsequence of X_n contains a subsequence weakly converging to a probability L . By assumption, X_n weakly converges to \mathbb{P}_X . By the characterization of the probability law by its characteristic function and by the unicity of weak limits, we have $L = \mathbb{P}_X$. Then, there exists a probability measure, $L_0 = \mathbb{P}_X$, such that each subsequence of X_n contains a subsequence weakly converging L_0 .

Step 2 : Let $f : \mathbb{R}^k \mapsto \mathbb{R}$ be a continuous and bounded function. Consider a subsequence $\mathbb{E}f(X_{n_j})$, $j \geq 1$, of the sequence of real numbers $\mathbb{E}f(X_n)$, $n \geq 1$.

This subsequence X_{n_j} , $j \geq 1$, contains a sequence $X_{n_{j_\ell}}$, $\ell \geq 1$, weakly converging to L_0 as $\ell \rightarrow +\infty$. Then $\mathbb{E}f(X_{n_{j_\ell}})$ converges to $\int f dL_0$. So $A = \int f dL_0$ is real number such that each subsequence of $\mathbb{E}f(X_n)$, $n \geq 1$ has a subsequence converging to A .

By Prohorov's Criterion Exercise 4 in Section 1 of Chapter 6, $\mathbb{E}f(X_n)$ converges to $\int f dL_0$. Then X_n weakly converges to $L_0 = \mathbb{P}_X$

CHAPTER 4

Specific Tools for Weak Convergence in \mathbb{R}

This chapter focuses on tools which are specific to convergence of sequences of real random variables. For such random variables, we may use Renyi's representations through uniform or exponential random variables, especially for sequence of independent and identically distributed random variables. Such representations use the generalized inverse functions on which concentrates the first section. Besides, in relation of Section 6 and Theorem 4.1 in Chapter 2, working on weak convergence in the same probability space may become a computations matter. This chapter gives the tools of such an orientation.

1. Generalized inverses of monotone functions

This theory is done for non-decreasing and right-continuous functions. It may be done for nonincreasing and left-continuous functions.

Sometimes, left or right continuity is not required (see Point 9 below).

Let F be a nondecreasing and right-continuous function from \mathbb{R} to \mathbb{R} . Let us define the generalized inverse of F as :

$$F^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}, u \in \mathbb{R}.$$

Because of the importance of this transformation for univariate extreme values theory, we are going to expose important facts of generalized inverses. Since we want them to be known by heart, we expose all of them before we provide their proofs.

A - List of most important properties of the generalized inverses.

Point (1). For any $u \in \mathbb{R}$ and for any $t \in \mathbb{R}$

(A)
$$F(F^{-1}(u)) \geq u$$

and

$$(B) \quad F^{-1}(F(x)) \leq x.$$

Point (2). For any $(u, t) \in \mathbb{R}^2$,

$$(A) \quad (F^{-1}(u) \leq t) \iff (u \leq F(t))$$

and

$$(B) \quad (F^{-1}(u) > t) \iff (u > F(t)).$$

Point (3). F^{-1} is nondecreasing and left-continuous.

Point (4). The weak convergence for non-decreasing distribution functions is available by itself and is defined still by Formula (9) above. Then we have the following implication.

$$(F_n \rightsquigarrow F) \Rightarrow (F_n^{-1} \rightsquigarrow F^{-1})$$

Point (5). Let us suppose that F_n and F are **distribution functions of real random variables** and that $F_n \rightsquigarrow F$. If F is **continuous**, that we have the uniform convergence

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0 \text{ as } n \rightarrow +\infty$$

Point (6). A distribution function F on \mathbb{R} has at most a countable points of discontinuity.

Point (7). Let \mathbb{P} be any probability measure on \mathbb{R} with support (a, b) , meaning that

$$a = \inf\{x, \mathbb{P}([-\infty, x]) > 0\} \text{ and } b = \inf\{x, \mathbb{P}([-\infty, x]) = 1\}.$$

which $\mathbb{P}((a, b)^c) = 0$. Then, for $0 < \varepsilon < 1$, there exists a finite number partition of (a, b) ,

$$a = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} = b$$

such that for $0 < i < k$,

$$\mathbb{P}(]t_i, t_{i+1}[) \leq \varepsilon.$$

We always can extend the bounds to

$$-\infty \leq t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} \leq +\infty$$

since $\mathbb{P}(]-\infty, a]) = 0$ and $\mathbb{P}(]b, +\infty[) = 0$.

Point (8) Let F and G be two distribution function both nonincreasing or both nondecreasing. If neither of them is degenerated, then there exist two continuity points of both F and G x_1 and x_2 such that $x_1 < x_2$ and

$$F(x_1) < F(x_2) \text{ and } G(x_1) < G(x_2).$$

Point (9) Let F be simply non-decreasing from \mathbb{R} to $[a, b]$ without assumption of left or right-continuity. Then for any $y \in]a, b[$,

$$F(F^{-1}(y) + 0) \leq y \leq F(F^{-1}(y) - 0),$$

where $F(\cdot +)$ and $F(\cdot -)$ respectively stand the right and the left limit at x .

If the function F is non-increasing, the generalized inverse is defined by

$$F^{-1}(y) = \inf\{x \in \mathbb{R}, F(x) \leq y\}, y \in (a, b).$$

and we have the formula for $x \in (a, b)$

$$F(F^{-1}(y) + 0) \leq y \leq F(F^{-1}(y) - 0)$$

B - Proofs of the points.

Proof of Point 1. Part (A). Set

$$A_u = \{x \in \mathbb{R}, F(x) \geq u\}, u \in \mathbb{R}.$$

Since $F^{-1}(u) = \inf A_u$, there exists a sequence $(x_n)_{n \geq 0} \in A_u$ such that

$$\begin{cases} F(x_n) & \geq & u \\ x_n & \downarrow & F^{-1}(u) \end{cases}$$

By right-continuity of F we have

$$F(F^{-1}(u)) \geq u.$$

This proves first Formula (A). As to the Formula (B), consider $x \in \mathbb{R}$ and set

$$F^{-1}(F(x)) = \inf A_{F(x)}.$$

Let us split $A_{F(x)}$ into

$$A_{F(x)} = [-\infty, x[\cap A_{F(x)} + [x, +\infty] \cap A_{F(x)}$$

$$=: A_{F(x)}(1) + A_{F(x)}(2).$$

By Fact 1 below, at the end of this subsection, we have

$$\inf A_{F(x)} = \min(\inf A_{F(x)}(1), \inf A_{F(x)}(2))$$

But

$$y \in A_{F(x)}(1) \implies y \leq x, \text{ then } \inf A_{F(x)}(1) \leq x$$

Next we obviously have

$$\inf A_{F(x)}(2) = \{y \geq x, F(y) \geq F(x)\} = x.$$

Thus

$$\inf A_{F(x)} \leq x.$$

That is :

$$F^{-1}(F(x)) = \inf A_{F(x)} \leq x.$$

This closes the proof of Point 1.

Proof of Point 2. It is obvious that each of Formulas (A) and (B) is derived from the other by taking complementary. So, we may only prove on them, say (B). Suppose $(u > F(t))$. By right-continuity of F at t , we can find ε such that

$$u > F(t + \varepsilon).$$

Now, for $x \in A_u$ we surely have

$$x > t + \varepsilon.$$

Otherwise, we would get,

$$x \leq t + \varepsilon \implies F(x) \leq F(t + \varepsilon) < u,$$

and this would lead to the conclusion $x \notin A_u$, which is in contradiction with the assumption. So, $x > t + \varepsilon$ for all $x \in A_u$. This implies that

$$\inf A_u = F^{-1}(u) \geq t + \varepsilon > t$$

We proved the direct sens of the first formula. To prove the indirect sense, consider $F^{-1}(u) > t$. Next, suppose that $u > F(t)$ does not hold. This implies that $F(t) \geq u$, which is in contradiction with $t \in A_u$ and next,

$$\inf A_u = F^{-1}(u) \leq t.$$

This is impossible. Then $u > F(t)$.

Proof of Point 3. We begin to establish that F^{-1} is non-decreasing. We have

$$\forall u \leq u', A_{u'} \leq A_u \implies \inf A_{u'} \leq \inf A_u.$$

This implies

$$F^{-1}(u') \geq F^{-1}(u)$$

Next, we have to prove that F^{-1} is left-continuous. Let $u \in \mathbb{R}$. We have for any $h \geq 0$,

$$F^{-1}(u - h) \leq F^{-1}(u).$$

Thus

$$\lim_{h \downarrow 0} F^{-1}(u - h) \leq F^{-1}(u).$$

Suppose that

$$\lim_{h \downarrow 0} F^{-1}(u - h) = \alpha < F^{-1}(u).$$

We can find $\varepsilon > 0$ such that $\alpha + \varepsilon < F^{-1}(u)$. Now, for all $h \geq 0$,

$$F^{-1}(u - h) < \alpha + \varepsilon.$$

By definition of the infimum, there exists x such that

$$F(x) \geq u - h \text{ and } F^{-1}(u - h) < \alpha + \varepsilon$$

By Formula (A) of Point 1 and Formula (B) of Point (2), we have

$$F^{-1}(u - h) < \alpha + \varepsilon \implies u - h \leq F(\alpha + \varepsilon).$$

Then we get as $h \downarrow 0$

$$u \leq F(\alpha + \varepsilon).$$

Since this is true for any $\varepsilon > 0$, we let $\varepsilon \downarrow 0$ to get

$$u \leq F(\alpha)$$

But, by Formula (B) of Point (2) and by using the hypothesis, we arrive at

$$(\alpha < F^1(u)) \Leftrightarrow (F^{-1}(u) > \alpha) \Leftrightarrow (u > F(\alpha)).$$

This is clearly a contradiction. We conclude that

$$\lim_{h \downarrow 0} F^{-1}(u - h) = F^{-1}(u).$$

And next, F^{-1} is left-continuous.

Proof of Point 4

Suppose that $F_n \rightsquigarrow F$. Let $y \in \mathbb{R}$ and let $\varepsilon > 0$. Since the number of discontinuity of F is at most countable, we can find a continuity point x of F in the open interval $(F^{-1}(y) - \varepsilon, F^{-1}(y))$. By Point 2, $(F^{-1}(y))$ is equivalent to $(F(x) < y)$. Since $x \in C(F)$, $F_n(x) \rightarrow F(x)$. Then for values of n large enough, we have $F_n(x) < y$ and then $x < F_n^{-1}(y)$. We get

$$F^{-1}(y) - \varepsilon \leq x < F_n^{-1}(y)$$

that is for any $\varepsilon > 0$,

$$F_n^{-1}(y) > F^{-1}(y) - \varepsilon.$$

We let $n \rightarrow \infty$ and ε to decrease to 0, and we get for any $y \in \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} F_n^{-1}(y) \geq F^{-1}(y).$$

Now let y be a continuity of F^{-1} . For any $y' > y$, we can find a continuity point x of F such that

$$(1.1) \quad F^{-1}(y') < x < F^{-1}(y') + \varepsilon.$$

By Point 1, $x > F^{-1}(y') \implies F(x) \geq F(F^{-1}(y')) \geq y'$. Then

$$y < y' \leq F(x).$$

Since $x \in C(F)$, we have $F_n(x) \rightarrow F(x)$, and then for large values of n , $y < F_n(x)$ and by Formula (A) of Point 2, $F_n^{-1}(y) \leq x$. By combining this with Formula (1.1), we get

$$F^{-1}(y') \geq x \geq F_n^{-1}(y).$$

Now let $n \rightarrow \infty$ to obtain

$$\limsup_{n \rightarrow \infty} F_n^{-1}(y) \leq F^{-1}(y').$$

Next, let $y' \downarrow y$, and get $F^{-1}(y') \downarrow F^{-1}(y)$ by continuity of F^{-1} en y . We arrive at

$$\limsup_{n \rightarrow \infty} F_n^{-1}(y) \leq F^{-1}(y) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(y).$$

We finally conclude that

$$F^{-1}(y) = \limsup_{n \rightarrow \infty} F_n^{-1}(y) = \liminf_{n \rightarrow \infty} F_n^{-1}(y).$$

Proof of Point 5. Since F is non-decreasing, x is a discontinuity point of F if and only if the discontinuity jump $F(x+) - F(x-)$ is positive. Denote by D the set of all discontinuity points of F , and for any $k \geq 1$, denote by D_k the set of discontinuity points such that $F(x+) - F(x-) > 1/k$ and by $D_{k,n}$ the set of discontinuity points in the interval $[-n, n]$ such that $F(x+) - F(x-) > 1/k$. We are going to show that $D_{k,n}$ is finite.

Let us suppose we can find m points x_1, \dots, x_m in $D_{k,n}$. Since F is non-decreasing, we may see that the sum of the discontinuity jumps is less than $F(n) - F(-n)$. You may make a simple drawing for $m = 3$ and project the jumps to the y-axis to see this easily. So

$$\sum_{1 \leq j \leq m} F(x_j+) - F(x_j-) \leq F(n) - F(-n).$$

Since each of these jumps exceeds $1/k$, we have

$$\sum_{1 \leq j \leq m} (1/k) \leq \sum_{1 \leq j \leq m} F(x_j+) - F(x_j-) \leq F(n) - F(-n).$$

$$m/k \leq F(n) - F(-n)$$

That is

$$m \leq k(F(n) - F(-n))$$

We conclude by saying that we cannot have more than $[k(F(n) - F(-n))]$ points in $D_{k,n}$, so $D_{k,n}$ is finite. Since

$$D = \bigcup_{n \geq 1} \bigcup_{k \geq 1} D(k, n),$$

we see that D is countable. This puts an end to the proof.

Proof of Point 6. Let $0 < \varepsilon < 1$. Let $F(t) = \mathbb{P}([-\infty, t])$. This is a distribution function such that $F(\infty) = 0$ and $F(+\infty) = 1$. Fix

$0 < \varepsilon < 1$. Set $k = [1/\varepsilon]$, where $[t]$ stands for the geatest integer less or equal to t . We then have

$$k\varepsilon \leq 1 \leq k\varepsilon + \varepsilon$$

and denote

$$s_i = i\varepsilon, \text{ for } i = 1, \dots, k \text{ and } s_{k+1} = 1.$$

Put

$$t_i = F^{-1}(s_i) = \inf\{u, G(u) \geq s_i\}.$$

By Point 1,

$$(1.2) \quad F(t_i) \geq s_i.$$

Next, for any $1 \leq i < k$,

$$F(t_{i+1}-) = \lim_{h \downarrow 0} F(t_{i+1} - h).$$

By definition of t_{i+1} , which the supremum of the values u such that $F(u) \geq (i+1)\varepsilon$, we surely have,

$$F(t_{i+1} - h) < (i+1)\varepsilon.$$

By letting $h \downarrow 0$, we get

$$(1.3) \quad F(t_{i+1}-) \leq (i+1)\varepsilon.$$

By putting together, (1.2) et (1.3), we have

$$\mathbb{P}(]t_i, t_{i+1}[) = F(t_{i+1}-) - F(t_i) \leq (i+1)\varepsilon - i\varepsilon = \varepsilon,$$

for $i = 1, \dots, k$. For $i = k$, we have $F(t_{k+1}) = 1$ and

$$\mathbb{P}(]t_k, t_{k+1}[) = 1 - F(t_k) \leq 1 - k\varepsilon \leq \varepsilon.$$

For $i = 0$, since $F(t_0) \geq 0$, we have

$$\mathbb{P}(]t_0, t_{i+1}[) = F(t_{i+1}-) - F(t_0) \leq F(t_{i+1}-) \leq \varepsilon$$

We just proved that $0 \leq i \leq k$,

$$\mathbb{P}(]t_i, t_{i+1}[) = F(t_{i+1}+) - F(t_i) \leq (i+1)\varepsilon - i\varepsilon = \varepsilon.$$

Proof of Point 7. We are going to apply Point 6. Let us consider the Lebesgues-Stieljes probabibiliy measure generated by F and characterized by

$$\mathbb{P}(]u, v]) = F(v) - F(u), u \leq v.$$

In particular, we have $\mathbb{P}(]-\infty, v]) = F(v) - F(u)$. Fix $\varepsilon > 0$ and consider a subdivision

$$-\infty = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} = +\infty$$

such that for any $0 \leq j \leq k$,

$$F(t_{j+1}-) - F(t_j) = \mathbb{P}_X([t_j, t_{j+1}[) \leq \varepsilon.$$

Now we want to prove the uniform convergence. Let x be one of the t_j 's. We have

$$F_n(x) - F(x) \leq \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|.$$

any other x is in one of the intervals $]t_j, t_{j+1}[$. Use the nondecreasingness of F and F_n to have

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(t_{j+1}-) - F(x) \\ &\leq F_n(t_{j+1}-) - F(t_{j+1}-) + F(t_{j+1}-) - F(x) \\ &\leq F_n(t_{j+1}-) - F(t_{j+1}-) + F(t_{j+1}-) - F(t_j) \\ &\leq \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)| + \varepsilon \end{aligned}$$

and

$$\begin{aligned} F(x) - F_n(x) &\leq F(x) - F_n(t_j) \\ &\leq F(x) - F(t_j) + F(t_j) - F_n(t_j) \\ &\leq F(t_{j+1}-) - F(t_j) + F(t_j) - F_n(t_j) \\ &\leq \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)| + \varepsilon \end{aligned}$$

At the arrival, we have for any point x different from the t_j ,

$$|F(x) - F_n(x)| \leq \max\left(\sup_{0 \leq j \leq k+1} |F_n(t_j-) - F(t_j-)|, \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|\right) + \varepsilon.$$

Then

$$\begin{aligned} \sup_{x \in R} |F_n(x) - F(x)| &= \|F_n - F\|_\infty \\ &\leq \max\left(\sup_{0 \leq j \leq k+1} |F_n(t_j-) - F(t_j-)|, \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|\right) \\ &\quad + \varepsilon. \end{aligned}$$

At this step, we have the more general conclusion. If for all real x , $F_n(x) \rightarrow F(x)$ and $F_n(x-) \rightarrow F(x-)$, then we may conclude that

$$\limsup_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq +\varepsilon,$$

for an arbitrary ε . Thus we have

$$\limsup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0.$$

To extend this conclusion to the case F is continuous and $F_n(x) \rightarrow F(x)$, we have to prove $F_n(x-) \rightarrow F(x)$ for any x .

To prove this, fix an arbitrary x and let $0 \leq h_p \downarrow 0$ as $p \uparrow +\infty$. We have for each n ,

$$F_n(x-) - F(x) \leq F_n(x) - F(x) \leq |F_n(x-) - F(x)|,$$

and

$$F(x) - F_n(x-) \leq F(x) - F_n(x - h_p) \leq |F(x) - F(x - h_p)| + |F(x - h_p) - F_n(x - h_p)|.$$

By combining these two points, we have

$$|F(x) - F_n(x-)| \leq \max(|F_n(x-) - F(x)|, |F(x) - F(x - h_p)| + |F(x - h_p) - F_n(x - h_p)|).$$

Now fix p and let $n \rightarrow +\infty$ to get

$$\limsup_{n \rightarrow +\infty} |F(x) - F_n(x-)| \leq |F(x) - F(x - h_p)|.$$

Finally, let $p \rightarrow +\infty$ to get the conclusion by continuity of F .

Proof of Point 8. F is degenerated if and only if it has a unique point of increase, say a , at which it presents a discontinuity jump. It is in the form : $F(x) = c_1$ for $x < a$ and $F(x) = c_2$ for $x \geq a$, with $c_1 < c_2$. So if F is non-degenerated, it has at least to two points of increase. Hence, we can find three continuity points of F : $a_1 < a_1 < a_3$ such that $F(a_1) < F(a_1) < F(a_3)$. If G is also non degenerated, we also find three continuity points of G : $b_1 < b_1 < b_3$ such that $F(b_1) < F(b_1) < F(b_3)$. We consider two cases.

Case 1. The intervals $[a_1, a_3]$ and $[b_1, b_3]$ are disjoint or have an intersection of one point. Suppose for example that $a_3 \leq b_1$. Take $x_1 = a_1$ and $x_2 = b_3$. We have

$$F(x_1) < F(a_3) \leq F(b_1) \leq F(b_3) = F(x_2)$$

and

$$G(x_1) \leq G(a_3) \leq G(b_1) < F(b_3) = G(x_2).$$

Case 2. The intervals $[a_1, a_3]$ and $[b_1, b_3]$ overlap at least on a non-empty open interval. Take t in the intersection. Surely we have $F(a_1) < F(t)$ or $F(t) < F(a_3)$. Otherwise, we would have $F(a_1) = F(a_3)$, which violates what is above. Similarly $G(b_1) < G(t)$ or $G(t) < G(a_3)$. Now, take $x_1 = \min(a_1, b_1)$ and $x_2 = \min(a_3, b_3)$.

If $F(a_1) < F(t)$, we have

$$F(x_1) \leq F(a_1) < F(t) \leq F(a_3) \leq F(x_2)$$

If $F(t) < F(a_3)$, we have

$$F(x_1) \leq F(a_1) \leq F(t) < F(a_3) \leq F(x_2)$$

We conclude that

$$F(x_1) < F(x_2)$$

We prove similarly that

$$G(x_1) < G(x_2)$$

By Point 5, we know that the discontinuity points of F and G are at most countable, we may adjust x_1 and x_2 to be continuity points of both F and G .

Proof of Point 9.

Let us begin by the first case where F is nondecreasing. By definition of the generalized inverse, we have for any $h > 0$,

$$F(F^{-1}(y) + h) \geq y$$

and

$$F(F^{-1}(y) - h) < y.$$

By letting h decrease to zero, we get

$$F(F^{-1}(y)-) \leq y \leq F(F^{-1}(y)+), \quad y \in (a, b).$$

Similarly, if F is non-increasing, we have for any $h > 0$

$$F(F^{-1}(y) + h) \leq y$$

and

$$F(F^{-1}(y) - h) > y.$$

By letting h decrease to zero, we get

$$F(F^{-1}(y)+) \leq y \leq F(F^{-1}(y)-).$$

Fact 1. Let A and B two disjoint subsets of \mathbb{R} . We have

$$\inf A \cup B = \min(\inf A, \inf B).$$

Indeed, clearly, $\inf A \cup B$ is less than $\inf A$ and less than $\inf B$, and then $\inf A \cup B \leq \min(\inf A, \inf B)$. Now suppose that we do not have the equality, that is

$$\inf A \cup B < \min(\inf A, \inf B).$$

There exists a sequence $(z_n)_{n \leq 0}$ of points of $A \cup B$ decreasing to $\inf(A \cup B)$. Surely for n large enough, z_n will be less than $\inf A$ and less than $\inf B$. And yet, it is either in A or in B . This is absurde. We conclude that we have the equality.

2. Applications of Generalized functions

The first application is the representation of any real random variable by a standard uniform random variable $U \sim \mathcal{U}(0, 1)$ associated with the distribution function $G(x) = 0$ for $x < 0$, $G(x) = x$ for $x \in (0, 1)$ and $G(x) = 1$ for $x > 0$. We have :

LEMMA 4. *Let F be a distribution function such that $F(-\infty) = 0$ and $F(+\infty) = 1$. Let $U \sim \mathcal{U}(0, 1)$, defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then $X = F^{-1}(U)$ has the distribution function F .*

Proof. We have by Formula (A) of Point 2 above that

$$\mathbb{P}(X \leq x) = P(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

A second application is this simple form of Skorohod Theorem.

THEOREM 12. *Let $F_n \rightsquigarrow F$, where F_n and F are distribution functions such that $F_n(-\infty) = 0$ and $F_n(+\infty) = 1$, for $n \geq 0$, $F(-\infty) = 0$ and $F(+\infty) = 1$. Then, there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a sequence of real random variables X_n and a random variable X such that for any $n \geq 0$, F_n is the distribution function of X_n , that is $F_n(\cdot) = \mathbb{P}(X_n \leq \cdot)$, and F is the distribution function of X , that is $F(\cdot) = \mathbb{P}(X \leq \cdot)$ and*

$$X_n \rightarrow X \text{ a.s. as } n \rightarrow +\infty.$$

Proof. Let us consider $([0, 1], \mathcal{B}([0, 1]), \lambda)$ where λ is the Lebesgue-measure on $[0, 1]$, which is a probability measure. Consider the identity function : $U : ([0, 1], \mathcal{B}([0, 1]), \lambda) \mapsto ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Then U follows a standard uniform law since for any $x \in (0, 1)$,

$$\lambda(U \leq x) = \lambda(U^{-1}([-\infty, x]))$$

where U^{-1} is the image-inverse of U and then $U^{-1}([-\infty, x]) =]-\infty, x]$. Thus

$$\lambda(U \leq x) = \lambda(U^{-1}([-\infty, x])) = \lambda([-\infty, x]) = x.$$

Consider $X_n = F_n^{-1}(U)$ and $X = F^{-1}(U)$. Then each F_n is the distribution function of X_n and F is the distribution function of X . Let us show that X_n converges to X almost-surely. By using Point 4 above, we have $F_n^{-1} \rightsquigarrow F^{-1}$. Then

$$\begin{aligned} 1 &\geq \lambda(X_n \rightarrow X) = \lambda(\{u \in [0, 1], X_n(u) \rightarrow X(u)\}) \\ &= \lambda(\{u \in [0, 1], F_n^{-1}(u) \rightarrow F^{-1}(u)\}) \\ &\geq \lambda(\{u \in [0, 1], u \text{ is a continuity point of } F\}) = 1, \end{aligned}$$

since the complement of $\{u \in [0, 1], u \text{ is a continuity point of } F\}$ is countable and countable sets are null-set with respect to the Lebesgue-measure.

3. Representation of Renyi for iid sequences of random Variables

This section is intended to provide representations of order statistics $X_{1,n} \leq \dots \leq X_{n,n}$, $n \geq 1$, of any n independent random variables X_n, \dots, X_n with common distribution function F in that of standard uniform or exponential independent random variables.

We remind again that in this section, all the random variables are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

We begin by recalling the density probability function of the order statistics from a density probability function h .

3.1. Density of the order statistics.

PROPOSITION 27. *Let Z_1, Z_2, \dots, Z_n be n independent copies of an absolutely continuous random variable Z of probability density function h , defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The associated order statistics $Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n,n}$ has the joint probability density function*

$$h_{(Z_{1,n}, \dots, Z_{n,n})}(z_1, \dots, z_n) = n! \prod_{i=1}^n h(z_i) 1_{(z_1 \leq \dots \leq z_n)}.$$

Proof. Suppose that the assumptions of the proposition holds. Let us find the joint density probability functions of r order statistics $Z_{n_1,n} \leq Z_{n_2,n} \leq \dots \leq Z_{n_r,n}$, with $1 \leq r \leq n$, $1 \leq n_1 < n_2 < \dots < n_r$.

Since Z is an absolutely continuous random variable, the observations are distinct almost surely and we have $Z_{n_1,n} < Z_{n_2,n} < \dots < Z_{n_r,n}$. Then for dz_i small enough and for $z_1 < z_2 < \dots < z_r$, the event

$$(Z_{n_i,n} \in]z_i - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)$$

occurs with $n_1 - 1$ observations of the sample Z_1, \dots, Z_n falling at left of z_1 , one point in $]z_1 - dz_1/2, z_1 + dz_1/2[$, $n_2 - n_1 - 1$ between $z_1 + dz_1/2$ and $z_2 - dz_1/2$, one point in $]z_2 - dz_2/2, z_1 + dz_2/2[$, etc., and $n - k_k$ points at right of z_r .

This is illustrated in Figure 1 for $r = 3$.

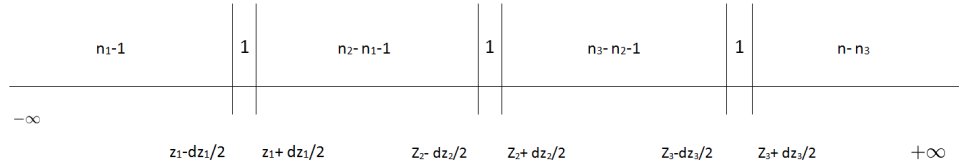


FIGURE 1. How are placed the observations with respect to $z_1 < \dots < z_r$

By definition, the probability density function $f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}$, whenever it exists, satisfies

$$(3.1) \quad \frac{\mathbb{P}(Z_{n_i,n} \in]z_1 - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)}{dz_1 \times \dots \times dz_r} = f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}(z_1, \dots, z_r)(1 + \varepsilon(dz_1, \dots, dz_r)),$$

where $\varepsilon(dz_1, \dots, dz_r) \rightarrow 0$ as each $\Delta_i \rightarrow 0$ ($1 \leq i \leq r$). Now, by using the independance Z_1, \dots, Z_n , $\mathbb{P}(Z_{n_i,n} \in]z_1 - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)$ is obtained as a multinomial probability. Using in addition the fact that h is the common probability density function of Z , we get

$$\begin{aligned} & \frac{\mathbb{P}(Z_{n_i,n} \in]z_1 - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)}{dz_1 \times \dots \times dz_r} \\ &= n! \times \frac{h(z_1)^{n_1-1}}{(n_1-1)!} \times \frac{(F(z_2) - F(z_1))^{n_2-n_1-1}}{(n_2-n_1-1)!} \\ &\times \dots \times \frac{(F(z_j) - F(z_{j-1}))^{n_j-n_{j-1}-1}}{((n_j-n_{j-1}-1)!) } \\ &\times \dots \times \frac{(F(z_r) - F(z_{r-1}))^{n_r-n_{r-1}-1}}{(n_r-n_{r-1}-1)!} \\ &\times \frac{(1 - F(z_r))^{n-n_r}}{(n-n_r)!} \\ &\times \prod \frac{\mathbb{P}(Z_{n_i,n} \in]z_i - dz_i/2, z_i + dz_i/2[)}{1! \Delta_i} \end{aligned}$$

The last factor in the latter product is

$$\prod_{i=1}^r h(z_i)(1 + dz_i).$$

By setting $n_0 = 0$ and $n_r = n + 1$ and for $-\infty = z_0 < z_1 < \dots < z_r < z_{r+1} = +\infty$,

$$f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}(z_1, \dots, z_r) = n! \prod_{j=1}^{r+1} \frac{h(z_j)(F(z_j) - F(z_{j-1}))^{n_j-n_{j-1}-1}}{(n_j-n_{j-1}-1)!},$$

we see that $f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}$ satisfies (3.1). Then we have the partial conclusion :

LEMMA 5. *Let Z_1, Z_2, \dots, Z_n be n independent copies of an absolutely continuous random variable Z of probability density function h and probability distribution function H , defined on the same probability*

space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $1 \leq r \leq n$, $1 \leq n_1 < n_2 < \dots < n_r$. Then the r order statistics $Z_{n_1, n} < Z_{n_2, n} < \dots < Z_{n_r, n}$ have the joint probability density function in (z_1, \dots, z_r) ,

$$(3.2) \quad n! \prod_{j=1}^{r+1} \frac{h(z_j)(F(z_j) - F(z_{j-1}))^{n_j - n_{j-1} - 1}}{(n_j - n_{j-1} - 1)!} 1_{(z_1 < \dots < z_r)},$$

with by convention $n_0 = 0$ and $n_r = n + 1$, $z_0 = -\infty = z_0$ and $z_{r+1} = +\infty$.

To finish the proof, let $r = n$ and $n_1 = 1, n_2 = 2, \dots, n_n = n$. Since the numbers $n_j - n_{j-1} - 1$ vanish in (3.2), It comes that $Z_{1, n} < Z_{2, n} < \dots < Z_{n, n}$ have the joint probability density

$$n! \prod_{j=1}^n h(z_j) 1_{(z_1 < \dots < z_r)},$$

Now, we are focusing on the relation between standard uniform and exponential order statistics.

PROPOSITION 28. *Let $n \geq 1$ be a fixed integer and $U_{1, n} \leq U_{2, n} \leq \dots \leq U_{n, n}$ be the order statistics associated with U_1, U_2, \dots, U_n , which are n independent random variables uniformly distributed on $(0, 1)$. Let $E_1, E_2, \dots, E_n, E_{n+1}$, $(n + 1)$ independent random variables following the standard exponential law, that is*

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(E_i \leq x) = (1 - e^{-x}) 1_{(x \geq 0)}, i = 1, \dots, n + 1.$$

Let $S_j = E_1 + \dots + E_j$, $1 \leq j \leq n + 1$. Then we have the following equality in distribution

$$(U_{1, n}, U_{2, n}, \dots, U_{n, n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right)$$

Proof. On one hand, by (27), the probability density function (pdf) of $U = (U_{1, n}, U_{2, n}, U_{n, n})$ is given by

$$\forall (u_1, \dots, u_n) \in \mathbb{R}^n, f_U(u_1, \dots, u_n) = n! 1_{(0 \leq u_1 \leq \dots \leq u_n \leq 1)}.$$

We are going to find the distribution of $Z_{n+1}^* = (S_1, S_2, \dots, S_n, S_{n+1})$ given $S_{n+1} = t$, $t > 0$. We have for $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$,

$$(3.3) \quad \begin{aligned} f_{Z_n^*}^{S_{n+1}=t}(y) &= \frac{f_{(Z_n^*, S_{n+1})}(y, t)}{f_{S_{n+1}}(t)} \\ &= \frac{f_{Z_{n+1}^*}(y, t)}{f_{S_{n+1}}(t)} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq t)}. \end{aligned}$$

But S_{n+1} follows a gamma law of parameters $n+1$ and 1, that is $S_{n+1} \sim \gamma(n+1, 1)$, and its probability density function is

$$(3.4) \quad f_{S_{n+1}}(t) = \frac{t^n e^{-t}}{\Gamma(n+1)} \mathbf{1}_{(t \geq 0)} = \frac{t^n}{n!} e^{-t} \mathbf{1}_{(t \geq 0)}.$$

The distribution function of (S_1, \dots, S_{n+1}) comes from the transformation

$$\begin{pmatrix} E_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ E_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ 0 & -1 & 1 & & & \\ \cdot & & \cdot & \cdot & & \\ \cdot & & & \cdot & \cdot & \\ \cdot & & & & -1 & 1 \end{pmatrix} \begin{pmatrix} S_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ S_{n+1} \end{pmatrix}$$

Let B the matrix on the formula above. The Jacobian determinant in absolute value is $|B| = 1$ and

$$B \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_{n+1} \end{pmatrix} = (y_1, y_2 - y_1, \dots, y_{n+1} - y_n).$$

Then, the density of (S_1, \dots, S_{n+1}) is then given by

$$\begin{aligned} f_{Z_{n+1}^*}(y_1, \dots, y_{n+1}) &= f_{(E_1, \dots, E_{n+1})}(B(y_1, \dots, y_{n+1})) \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})} \\ &= \prod_{i=1}^{n+1} e^{-(y_i - y_{i-1})} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})} \\ &= e^{-y_{n+1}} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})}. \end{aligned}$$

where convention $y_0 = 0$ by convention. Going back to (3.3) and (3.4), we get, with $y = (y_1, y_2, \dots, y_n)$,

$$(3.5) \quad f_{Z_n^*}^{S_{n+1}=t}(y) = \frac{n!}{t^n} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq t)}.$$

Now, for $u = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$,

$$f_{\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)}^{S_{n+1}=t}(u) = f_{\left(\frac{S_1}{t}, \dots, \frac{S_n}{t}\right)}^{S_{n+1}=t}(u_1, u_2, \dots, u_n).$$

This density probability function is obtained from (3.5) by the transform

$$(y_1, y_2, \dots, y_n) = t(u_1, u_2, \dots, u_n) \iff (u_1, u_2, \dots, u_n) = \frac{1}{t}(y_1, y_2, \dots, y_n)$$

with jacobian determinant t^n . Then

$$\begin{aligned} f_{\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)}^{S_{n+1}=t}(u_1, u_2, \dots, u_n) &= f_{Z_n^{S_{n+1}=t}}^{S_{n+1}=t}(t(u_1, u_2, \dots, u_n)) t^n \mathbf{1}_{(0 \leq tu_1 \leq \dots \leq tu_n \leq t)} \\ &= n! \mathbf{1}_{(0 \leq u_1 \leq \dots \leq u_n \leq 1)}. \end{aligned}$$

This is exactly (3.3). Then the conditional distribution of $Z = \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)$ given $S_{n+1} = t$ does not depend on t . So, its conditional distribution is also its unconditional distribution function and then

$$\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)$$

has the same law as $U = (U_{1,n}, U_{2,n}, \dots, U_{n,n})$ and it is independent of S_{n+1} . This puts an end to the proof.

We formalize the last conclusion in

LEMMA 6. *Let $E_1, E_2, \dots, E_n, E_{n+1}, n \geq 1$ be independent standard exponential random variables defined on the same probability space. Let $S_i = E_1 + E_2 + \dots + E_i, 1 \leq i \leq n+1$, then*

$$\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)$$

is independent of S_{n+1} .

The latter proposition exposed representations of order statistics of standard uniform random variables into that of standard exponential random variables. The following proposition reverses the situation.

PROPOSITION 29. *Assume the notations of Proposition 28 hold. Then for any $n \geq 1$,*

$$(-\log U_{1,n}, \dots, -\log U_{n,n}) \stackrel{d}{=} (E_{1,n}, \dots, E_{n,n}),$$

where $E_{1,n} \leq \dots \leq E_{n,n}$ are the order statistics of E_1, E_2, \dots, E_n , which are n independent and exponentially distributed with intensity one.

Proof. By Proposition 27, the *pdf* of $E_{1,n} \leq \dots \leq E_{n,n}$ is

$$(3.6) \quad f_Z(z) = n! e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)}, \quad z = (z_1, \dots, z_n) \in \mathbb{R}^n,$$

where $Z = (E_{1,n}, \dots, E_{n,n})$. The distribution of $Z^* = (-\log U_{1,n}, \dots, -\log U_{n,n})$ comes from that of $U = (U_{1,n}, \dots, U_{n,n})$ by the diffeomorphism $(z_1, \dots, z_n) = (-\log u_1, \dots, -\log u_n)$ which preserves the order of the arguments and has a jacobian determinant in absolute value equal to

$$\begin{aligned} \left| \frac{\partial U_i}{\partial z_j} \right| &= \left| \frac{\partial e^{-z_i}}{\partial z_j} \right| = |\text{diag}(-e^{-z_1}, \dots, -e^{-z_n})| \\ &= e^{-\sum_{i=1}^n z_i}. \end{aligned}$$

Then, the *pdf* of Z^* is then

$$\begin{aligned} f_{Z^*}(z_1, \dots, z_n) &= f_U(-e^{-z_1}, \dots, -e^{-z_n}) e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)} \\ &= n! e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)}. \end{aligned}$$

This *pdf* is that of $(E_{1,n}, \dots, E_{n,n})$ by (3.6). The proof ends here.

Let us give another version of the previous result. It is clear that for any standard uniform random variable U , we have $U \stackrel{d}{=} 1 - U$. Then for any $n \geq 1$, we have

$$(U_{1,n}, \dots, U_{n,n}) \stackrel{d}{=} (1 - U_{1,n}, \dots, 1 - U_{n,n}).$$

The equality in distribution in Proposition 29 becomes : for any $n \geq 1$,

$$(-\log(1 - U_{n,n}), \dots, -\log(1 - U_{1,n})) \stackrel{d}{=} (E_{1,n}, \dots, E_{n,n})$$

Let us go further and denote

$$\alpha_{i,n} = -\log(1 - U_{i,n}), \quad 1 \leq i \leq n.$$

Consider the transformation for $n \geq 1$,

$$\begin{pmatrix} n\alpha_{1,n} \\ (n-1)(\alpha_{2,n} - \alpha_{1,n}) \\ \vdots \\ (n-i+1)(\alpha_{i,n} - \alpha_{i-1,n}) \\ \vdots \\ 1(\alpha_{n,n} - \alpha_{n-1,n}) \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_i \\ \vdots \\ V_n \end{pmatrix}.$$

We have

$$\begin{pmatrix} \alpha_{1,n} \\ \alpha_{2,n} \\ \vdots \\ \vdots \\ \vdots \\ \alpha_{n,n} \end{pmatrix} = \begin{pmatrix} V_1/n \\ V_1/n + V_2/(n-1) \\ \vdots \\ \vdots \\ \vdots \\ V_1/n + V_2/(n-1) + \dots + V_{n-1}/2 + V_1/1 \end{pmatrix}.$$

The density probability function of (V_1, \dots, V_n) is given by

$$f_V(v_1, \dots, v_n) = f_{(\alpha_{1,n}, \dots, \alpha_{n,n})}(v_1/n, v_1/n + v_2/(n-1), \dots, v_1/n + v_2/(n-1) + \dots + v_n) \times |J(v)| \times \mathbf{1}_{D_V}(v).$$

The Jacobian determinant in absolute value of this transform is

$$|J(v)| = \frac{1}{n!}$$

and the domain of V is

$$D_V = \mathbb{R}_+^n.$$

We conclude by using (3.6) which gives the joint pdf of $(\alpha_{1,n}, \dots, \alpha_{n,n})$, and by denoting $s_i = v_1/n + v_2/(n-1) + \dots + v_i/(n-i+1)$, $i = 1, \dots, n$.

We get

$$\begin{aligned} f_V(v_1, \dots, v_n) &= \frac{1}{n!} \times n! e^{-\sum_{i=1}^n s_i} \mathbf{1}_{(v_1 \geq 0, \dots, v_n \geq 0)} \\ &= e^{-\sum_{i=1}^n s_i} \mathbf{1}_{(v_1 \geq 0, \dots, v_n \geq 0)}. \end{aligned}$$

We may check that $s_1 + \dots + s_n = v_1 + \dots + v_n$. We arrive at

$$f_V(v_1, \dots, v_n) = \prod_{i=1}^n e^{-v_i} \mathbf{1}_{(v_i \geq 0)}.$$

This says that (V_1, \dots, V_n) has independent standard exponential coordinates. We summarize our finding in :

PROPOSITION 30. *Let $\alpha_{i,n} = -\log(1 - U_{i,n})$, $i = 1, \dots, n$. Then the random variables $n\alpha_{1,n}$, $(n-1)(\alpha_{2,n} - \alpha_{1,n})$, \dots , $(n-i+1)(\alpha_{i,n} - \alpha_{i-1,n})$, \dots , $(\alpha_{n,n} - \alpha_{n-1,n})$ are independent standard exponential random variables.*

Let us do more and put for any $1 \leq i \leq n$,

$$(n-i+1)(\alpha_{i,n} - \alpha_{i-1,n}) = (n-i+1) \log \left(\frac{1 - U_{i-1,n}}{1 - U_{i,n}} \right).$$

By our previous results we have that the random variables

$$E_{n-i+1}^* = (n - i + 1) (\alpha_{i,n} - \alpha_{i-1,n}) = \log \left(\frac{1 - U_{n-i,n}}{1 - U_{n-i+1,n}} \right)^{(n-i+1)}$$

are independent and standard exponential random variables. We may and do change $U_{n-i,n}$ to $U_{i+1,n}$ to arrive at this celebrated representation.

PROPOSITION 31. (*Malmquist representation*). Let U_1, U_2, \dots, U_n be standard uniform random variables for $n \geq 1$. Let $0 \leq U_{1,n} < U_{2,n} < \dots < U_{n,n} \leq 1$ be their associated order statistics. Then the random variables

$$\log \left(\frac{U_{i+1,n}}{U_{i,n}} \right)^i, i = 1, \dots, n$$

are independent standard exponential random variables.

The functional Empirical Process As a General Tools in Asymptotic Statistics

1. Using the small o's and the big O's

In this chapter, we will show how combine all the concepts we have studied so far to get a yet simple but powerful tools that may be systematically used to find asymptotic normal laws in a great variety of problems, even in current research problems. We will first study the manipulations of the $o_{\mathbb{P}}$ and the $O_{\mathbb{P}}$ symbols concerning limits in probability. Next, we present the functional empirical process that is used here only in the frame of the finite distributions case. Next, we will give some cases as illustrations.

It is important to notice for once that the method given are valid for sequences of random variables and limit random variables defined of the same probability space. In consequence, we treat sequences of random variables $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}, (Z_n)_{n \geq 1}, \dots$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^k, k \geq 1$ and $(a_n)_{n \geq 1}, (b_n)_{n \geq 1}, (c_n)_{n \geq 1}$ are positive random numbers.

2. Stochastic o's and O's

I - Big O's and small o's almost surely.

DEFINITIONS.

(a) The sequence of real random variables $(X_n)_{n \geq 1}$ is said to be an o (read the name of the letter o) of a_n almost surely as $n \rightarrow +\infty$, denoted by

$$X_n = o(a_n), a.s. \text{ as } n \rightarrow +\infty,$$

if and onl if

$$(2.1) \quad \lim_{n \rightarrow +\infty} X_n/a_n = 0 \text{ a.s.}$$

(b) The sequence of real random variables $(X_n)_{n \geq 1}$ is said to be a big O of a_n almost surely as $n \rightarrow +\infty$, denoted by

$$X_n = O(a_n), a.s. \text{ as } n \rightarrow +\infty,$$

if and only if the sequence $\{|X_n|/a_n, n \geq 1\}$ is almost surely bounded, that is

$$(2.2) \quad \lim_{n \rightarrow +\infty} \sup |X_n|/a_n < +\infty, a.s.$$

BE CAREFUL. The equality signs used (2.3) and (2.2) are to be read in one direction only in the sense : the left member is a small o of a_n or a big O of a_n . Do not reverse the equality from left to right. For example, if X_n is an $o(n)$, it is also an $o(n^2)$ and we may write $o(n) = o(n^2)$ *a.s.* but you cannot write $o(n^2) = o(n)$ *a.s.* An example : $X_n = n^{3/2}$ is an $o(n^2)$ but is not an $o(n)$. This remark will extend to the notations of small o's and big O's in probability to be defined below.

Particular cases concerning the constants. If $a_n = C > 0$ for any $n \geq 1$, denoted $a_n \equiv C$, we have :

(i) $X_n = O(C)$ *a.s.* if and only if X_n/C is bounded *a.s.* if and only if X_n is bounded *a.s.* and we write

$$X_n = O(1) \text{ a.s.}$$

(ii) $X_n = o(C)$ *a.s.* if and only if $X_n/C \rightarrow 0$ *a.s.* if and only if $X_n/C \rightarrow 0$ *a.s.* and we write

$$X_n = o(1) \text{ a.s.}$$

(iii) For any constant $C > 0$, we may write $C = O(1)$.

PROPERTIES.

The properties are very numerous and the user has often to check new ones depending of his undergoing work. But a few of them must be known and ready to be used. Let us list them in three groups.

Group A. Properties of small o's.

(1) $o(a_n)o(b_n) = o(a_nb_n)$ *a.s.*

(2) (1) $o(o(a_n)) = o(a_n)$ *a.s.*

(3) If $b_n \geq a_n$ for all $n \geq 1$, $o(a_n) = o(b_n)$ *a.s.*

(4) $o(a_n) + o(a_n) = o(a_n)$ *a.s.*

(5) $o(a_n) + o(b_n) = o(a_n + b_n)$ *a.s.* and $o(a_n) + o(b_n) = o(a_n \vee b_n)$ *a.s.* where $a_n \vee b_n = \max(a_n, b_n)$.

(6) $o(a_n) = a_no(1)$ *a.s.* and $a_no(1) = o(a_n)$ *a.s.*

PROOFS. Each of these properties is quickly proved in :

(1) If $X_n = o(a_n)$ and $Y_n = o(b_n)$, then

$$\lim_{n \rightarrow +\infty} \frac{|X_n Y_n|}{a_n b_n} = \lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} \times \lim_{n \rightarrow +\infty} \frac{|Y_n|}{b_n} = 0 \text{ a.s.}$$

and then $X_n Y_n = o(a_n b_n)$ *a.s.*

(2) If $Y_n = o(a_n)$, *a.s.* and $X_n = o(Y_n)$, *a.s.*,

$$\lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} = \lim_{n \rightarrow +\infty} \left| \frac{X_n}{Y_n} \right| \times \frac{|Y_n|}{a_n} = \lim_{n \rightarrow +\infty} \left| \frac{X_n}{Y_n} \right| \times \lim_{n \rightarrow +\infty} \frac{|Y_n|}{a_n} = 0,$$

that is $X_n = o(a_n)$ *a.s.*

(3) If $X_n = o(a_n)$ and $b_n \geq a_n$ for all $n \geq 1$, then

$$0 \leq \lim_{n \rightarrow +\infty} \sup \frac{|X_n|}{b_n} = \lim_{n \rightarrow +\infty} \sup \frac{|X_n|}{a_n} \frac{a_n}{b_n} \leq \lim_{n \rightarrow +\infty} \sup \frac{|X_n|}{a_n} = 0 \text{ a.s.}$$

and $|X_n/b_n| \rightarrow 0$ *a.s.*, that is $X_n = o(b_n)$, *a.s.*

(4) If $X_n = o(a_n)$ and $Y_n = o(a_n)$, then

$$\lim_{n \rightarrow +\infty} \sup \frac{|X_n + Y_n|}{a_n} \leq \lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} + \lim_{n \rightarrow +\infty} \frac{|Y_n|}{a_n} = 0 \text{ a.s.}$$

and then $X_n + Y_n = o(a_n)$ *a.s.*

(5) To prove that $o(a_n) + o(b_n) = o(a_n + b_n)$ *a.s.*, use Point (3) to see that $o(a_n) = o(a_n + b_n)$ *a.s.* since $a_n + b_n \geq a_n$ for all $n \geq 1$ and as well $o(b_n) = o(a_n + b_n)$ *a.s.* and then use Point (3) to conclude. We prove that $o(a_n) + o(b_n) = o(a_n \vee b_n)$ *a.s.* in the very same manner.

(6) This is a simple rephrasing the definition.

Group B. Properties of big o's.

(1) $O(a_n)O(b_n) = O(a_nb_n)$ *a.s.*

(2) $O(O(a_n)) = O(a_n)$ *a.s.*

(3) If $b_n \geq a_n$ for all $n \geq 1$, $O(a_n) = O(b_n)$ *a.s.*

(4) $O(a_n) + O(a_n) = O(a_n)$ *a.s.*

(5) $O(a_n) + O(b_n) = O(a_n + b_n)$ *a.s.* and $O(a_n) + O(b_n) = O(a_n \vee b_n)$ *a.s.* where $a_n \vee b_n = \max(a_n, b_n)$.

(6) $O(a_n) = a_n O(1)$ *a.s.*, and $a_n O(1) = O(a_n)$ *a.s.*

PROOFS. These properties are proved exactly as those of **Group A**, where superior limits are used at the place of limits.

Group C. Properties of combinations of small o's and big O's.

(1) $o(a_n)O(b_n) = o(a_nb_n)$ *a.s.*

(2) $o(O(a_n)) = o(a_n)$ *a.s.* and $O(o(a_n)) = o(a_n)$, *a.s.*

(3a) If $a_n = O(b_n)$, *a.s.*, then $o(a_n) + O(b_n) = O(b_n)$ *a.s.*

(3b) If $b_n = O(a_n)$, *a.s.*, then $o(a_n) + O(b_n) = O(a_n)$ *a.s.*

(3c) If $b_n = o(a_n)$, *a.s.*, then $o(a_n) + O(b_n) = o(a_n)$ *a.s.*

(4) $(1 + o(a_n))^{-1} - 1 = o(a_n)$, *a.s.*

PROOFS.

(1) If $X_n = o(a_n)$ and $Y_n = O(b_n)$, then

$$\lim_{n \rightarrow +\infty} \sup \frac{|Y_n|}{b_n} = C < +\infty \text{ a.s.}$$

and

$$\begin{aligned} \lim_{n \rightarrow +\infty} \sup \frac{|X_n Y_n|}{a_n b_n} &= \lim_{n \rightarrow +\infty} \left(\left| \frac{X_n}{a_n} \right| \times \frac{|Y_n|}{b_n} \right) = \lim_{n \rightarrow +\infty} \sup \left| \frac{X_n}{a_n} \right| \times \lim_{n \rightarrow +\infty} \sup \frac{|Y_n|}{b_n} \\ &\leq C \lim_{n \rightarrow +\infty} \sup \left| \frac{X_n}{a_n} \right| = 0 \text{ a.s.} \end{aligned}$$

(2) Use Points (6) of Groups A and B to say

$$o(O(a_n)) = o(1) \times O(a_n) = a_n \times o(1) \times O(1) = a_n \times o(1) = o(a_n)$$

and

$$O(o(a_n)) = o(a_n)O(1) = a_n \times o(1) \times O(1) = a_n \times o(1) = o(a_n).$$

(3a-b-c) These three points are proved in similar ways. Let us give the details of (3b) for example. Let $X_n = o(a_n)$ and $Y_n = O(b_n)$ and $b_n = O(a_n)$. Then

$$\begin{aligned} o(a_n) + O(b_n) &= o(a_n) + O(O(a_n)) = o(a_n) + O(a_n) \\ &= a_n(o(1) + O(1)) = a_n \times O(1) = O(a_n). \end{aligned}$$

(4) We have

$$\begin{aligned} (1 + o(a_n))^{-1} - 1 &= \frac{o(a_n)}{1 + o(a_n)} = \frac{o(a_n)}{1 + o(a_n)} = o(a_n)O(1) \\ &= a_n o(1)O(1) = a_n o(1) = a_n o(a_n). \end{aligned}$$

II - Big O 's and small o 's in probability.**DEFINITIONS.**

(a) The sequence of real random variables $(X_n)_{n \geq 1}$ is said to be an o (read the name of the letter o) of a_n in probability as $n \rightarrow +\infty$, denoted by

$$X_n = o_{\mathbb{P}}(a_n), \text{ as } n \rightarrow +\infty,$$

if and only if

$$(2.3) \quad \lim_{n \rightarrow +\infty} X_n/a_n = 0 \text{ in probability,}$$

that is for any $\lambda > 0$

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0.$$

(b) The sequence of real random variables $(X_n)_{n \geq 1}$ is said to be a big O of a_n in probability as $n \rightarrow +\infty$, denoted by

$$X_n = O_{\mathbb{P}}(a_n), \text{ as } n \rightarrow +\infty,$$

if and only if the sequence $\{|X_n|/a_n, n \geq 1\}$ is bounded in probability, that is : For any $\varepsilon > 0$, there exists a constant $\lambda > 0$, such that

$$(2.4) \quad \inf_{n \geq 1} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon$$

which is equivalent to

$$(2.5) \quad \liminf_{\lambda \uparrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0.$$

Before we go further, let us prove this :

LEMMA 7. *Each of (2.4) and (2.5) is equivalent to : For any $\varepsilon > 0$, there exists an integer $N \geq 1$ a constant $\lambda > 0$, such that*

$$(2.6) \quad \inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon.$$

PROOFS. To prove that (2.4) and (2.6) are equivalent, it will be enough to show that (2.6) \implies (2.4) since the reverse implication is obvious. Suppose that (2.6), that is, for $\varepsilon > 0$, there exist $N \geq 1$ and a real number $\lambda_0 > 0$ such that

$$(2.7) \quad \forall (n \geq N), \mathbb{P}(|X_n/a_n| \leq \lambda_0) \geq 1 - \varepsilon.$$

If $N = 1$, then (2.4) holds. If not, we have for $j \in \{1, \dots, N-1\}$ fixed, $(|X_j/a_j| \leq \lambda) \uparrow \Omega$ as $\lambda \uparrow +\infty$. So by the Monotone Convergence Theorem, there exists for each $j \in \{1, \dots, N-1\}$ a real number $\lambda_j > 0$ such that $\mathbb{P}(|X_j/a_j| \leq \lambda_j) > 1 - \varepsilon$. We take $\lambda = \max(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ and get

$$\forall (n \geq 1), \mathbb{P}(|X_n/a_n| \leq \lambda) \geq 1 - \varepsilon,$$

which is (2.4). Now, let us prove that (2.5) \iff (2.6). First (2.5) means

$$\lim_{\lambda \uparrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0,$$

since $\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n)$ is nonincreasing as $\lambda \uparrow +\infty$ on $[0, 1]$.

We get for any $\varepsilon > 0$, there exists a real number $\lambda > 0$ such that

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = \lim_{N \uparrow +\infty} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) \leq \varepsilon/2.$$

Then for some $N > 0$,

$$\sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) \leq \varepsilon,$$

that is

$$\inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon,$$

which is (2.6). In turn, a rephrasing of this gives : for any $\varepsilon > 0$, there exists $N_0 > 0$ and a real number $\lambda_0 > 0$ such that

$$(2.8) \quad \inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda_0 a_n) \geq 1 - \varepsilon,$$

that is

$$\sup_{n \geq N} \mathbb{P}(|X_n| > \lambda_0 a_n) < \varepsilon,$$

which leads to

$$\inf_{N \geq 1} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda_0 a_n) < \varepsilon,$$

and next

$$\inf_{\lambda > 0} \inf_{N \geq 1} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) < \varepsilon,$$

which is (2.5).

COMMENTARIES, NOTATIONS AND SOMME LEMMAS.

(a) From Chapter 3, an $O_{\mathbb{P}}(1)$ is simply a tight sequence of random variables. From Theorem 8 of Chapter 3, we have that any sequence $X_n = O_{\mathbb{P}}(a_n)$ contains a subsequence $(X_{n_k})_{k \geq 1}$ such that $(X_{n_k}/a_{n_k})_{k \geq 1}$ weakly converges in \mathbb{R} .

(b) It may be convenient to rephrase (2.4) into the following sentence.

For any $\varepsilon > 0$ there exists a real number $\lambda > 0$ such that $|X_n| \leq \lambda a_n$ With Probability At Least Equal to $1 - \varepsilon$ for all $n \geq 1$.

By using the complementary events, we will say : for any $\varepsilon > 0$ there exists a real number $\lambda > 0$ such that $|X_n| > \lambda a_n$ With Probability At

Most Equal to ε for all $n \geq 1$.

With Probability At Least Equal to $1 - \varepsilon$ will be abbreviated by $WPAL(1 - \varepsilon)$.

As well, $WPAME(varepsilon)$ is an abbreviation of With Probability At Most Equal to ε .

For lengthy demonstrations, using types of sentences described above may be handy.

Now, we may give some important properties of the small o 's and the big O 's in probability.

We will need two other lemmas.

LEMMA 8. *We have the following properties*

(a) *If X_n is a sequence of k -random vectors converging in probability to k -random vector X , then $\|X_n\| = O_{\mathbb{P}}(1)$.*

(b) *If X_n is a sequence of random vectors with values in the metric space (S, d) converging in probability to a constant $C \in S$ and if g is a measurable mapping from (S, d) to an other metric space (E, r) . Then if g is continuous at C , then $g(X_n)$ converges in probability to C .*

(c) *Consider a sequence of k -random vectors $(X_n)_{n \geq 1}$ converging to zero in probability. Let $R(x)$ be real function of $x \in \mathbb{R}^k$ continuous at zero and such that $R(0) = 0$. If $R(x) = o(\|x\|^p)$ as $x \rightarrow 0$ for some $p > 0$, then $R(X_n) = o_{\mathbb{P}}(\|X_n\|^p)$. If $R(x) = O(\|x\|^p)$ as $x \rightarrow 0$ for some $p > 0$, then $R(X_n) = O_{\mathbb{P}}(\|X_n\|^p)$.*

Proof.

Proof of Point (a). If $X_n \rightarrow_{\mathbb{P}} X$, then by Proposition 11, $X_n \rightarrow_w X$ and by the continuous mapping Theorem 6 of Chapter 2 $\|X_n\| \rightarrow_w \|X\|$. Then by Theorem 4, we have for any continuity point of $F_{\|X\|}(\lambda) = P(\|X\| \leq \lambda)$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\|X_n\| > \lambda) = \mathbb{P}(\|X\| > \lambda) = F_{\|X\|}(\lambda).$$

Since the set of discontinuity points of $F_{\|X\|}$ is at most countable (see Point 6 of Chapter 4). Then apply the formula above for $\lambda \rightarrow +\infty$

while λ are continuity points. Since $1 - F_{\|X\|}(\lambda) \rightarrow 0$ as $\lambda \rightarrow +\infty$, then for any $\varepsilon > 0$, we are able to pick one value of $\lambda(\varepsilon)$ which is a continuity point of $F_{\|X\|}$ satisfying $1 - F_{\|X\|}(\lambda) < \varepsilon$. For any $\varepsilon > 0$, we have found $\lambda(\varepsilon) > 0$ such that

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda(\varepsilon)) \leq \varepsilon,$$

which implies

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) = 0.$$

Point (a) is proved.

Proof of Point (b). Assume the notations of this point and suppose that g is continuous at C . Let $\varepsilon > 0$. By the continuity of g at C , there exists $\eta > 0$ such that

$$d(x, C) < \eta \implies r(g(x), g(C)) < \varepsilon/2.$$

Now

$$\begin{aligned} \mathbb{P}(r(g(X_n), g(C)) > \varepsilon) &= \mathbb{P}(\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) \geq \eta\}) \\ &\quad + \mathbb{P}(\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) < \eta\}) \\ &\leq \mathbb{P}(d(X_n, C) \geq \eta), \end{aligned}$$

since $(\{r(g(X_n), g(C)) > \lambda\} \cap \{d(X_n, C) < \eta\}) \subset (\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) < \eta\} \cap \{r(g(X_n), g(C)) < \varepsilon/2\}) = \emptyset$. Then, since $X_n \rightarrow_{\mathbb{P}} C$, we have

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(r(g(X_n), g(C)) > \varepsilon) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, C) \geq \eta) = 0.$$

So Point (b) is true.

Proof of Point (c-1). Let $R(x) = o(\|x\|^p)$ as $x \rightarrow 0$. Then $g(x) = |R(x)|/\|x\|^p \rightarrow 0$ as $x \rightarrow 0$. By continuity of g at zero and by Point (a), $g(X_n) = |R(X_n)|/\|X_n\|^p \rightarrow_{\mathbb{P}} 0$ as $n \rightarrow +\infty$ whenever $X_n \rightarrow_{\mathbb{P}} X$ as $n \rightarrow +\infty$. Hence $R(X_n) = o_{\mathbb{P}}(\|X_n\|^p)$.

Proof of Point (c-2). Let $R(x) = O(\|x\|^p)$ as $x \rightarrow 0$. Then for any $\varepsilon > 0$, there exist $\eta > 0$ and $C > 0$ such $|R(x)|/\|x\|^p \leq C$ for all $\|x\| < \eta$. Then for $\lambda > C$,

$$\begin{aligned}
\mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) &= \mathbb{P}(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| \geq \eta\}) \\
&+ \mathbb{P}(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\}) \\
&\leq \mathbb{P}(\|X_n\| \geq \eta),
\end{aligned}$$

since $(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\}) \subset (\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\} \cap \{|R(X_n)| / \|X_n\|^p < C\}) = \emptyset$. Then for all $\lambda > C$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) = 0$$

and then

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) = 0.$$

We conclude that $|R(X_n)| / \|X_n\|^p = O_{\mathbb{P}}(1)$.

In some situations, we would be able to work with convergence in probability while we are not sure of measurability of some sequences. For example, using the Mean Value Theorem with real random sequences X_n and Y_n and real function g of class C^1 may lead to this kind of formula

$$g(X_n) - g(Y_n) = (X_n - Y_n)g'(Z_n),$$

with $\min(X_n, Y_n) \leq Z_n \leq \max(X_n, Y_n)$. Here, we know that $g'(Z_n)$ is measurable but we do not know if Z_n is. In such a situation, we may need the notion of outer probability.

DEFINITION 4. *Let $(u_n)_{n \geq 1}$ be a sequence of real-valued application defined on Ω . It has a measurable covering sequence if and only if there exists a sequence of nonnegative real random variables $(v_n)_{n \geq 1}$ defined on (Ω, \mathcal{A}) such that*

$$\forall (n \geq 1), u_n \leq v_n.$$

Next, $(u_n)_{n \geq 1}$ converge to a real-valued application u defined on Ω , as $n \rightarrow +\infty$, if and only if the sequence $(u_n - u)_{n \geq 1}$ has a measurable covering sequence which converges to zero in probability, and we denote

$$u_n \rightarrow_{\mathbb{P}^*} u \text{ as } n \rightarrow +\infty$$

We are going to see that the result of Point (b) of the lemma above still holds for convergence in outer probability in the special case of \mathbb{R} .

LEMMA 9. *Let X_n is a sequence of real-valued applications defined on Ω converging in outer probability to $c \in \mathbb{R}$. Let g is a real-valued function defined on \mathbb{R} , continuous at c and such that for each $n \geq 1$, $g(X_n)$ is measurable. Then $g(X_n)$ converges in probability to $g(c)$*

Proof. Assume the notations of the lemma. Let Y_n be a sequence of random variables such that $|X_n - c| \leq Y_n$ for all $n \geq 1$ and $Y_n \rightarrow 0$ in probability.

Now, by the continuity of g at c , there exists $\eta > 0$ such that

$$|x - c| < \eta \implies |g(x) - g(c)| < \varepsilon/2.$$

Next

$$\begin{aligned} \mathbb{P}(|g(X_n) - g(c)| > \varepsilon) &= \mathbb{P}(\{|g(X_n) - g(c)| > \varepsilon\} \cap \{Y_n \geq \eta\}) + \mathbb{P}(\{|g(X_n) - g(c)| > \varepsilon\} \cap \{Y_n < \eta\}) \\ &\leq \mathbb{P}(\{Y_n \geq \eta\}) \end{aligned}$$

since $\{|g(X_n) - g(c)| > \varepsilon\} \cap \{Y_n < \eta\} = \emptyset$. The reason on this is that on $\{|g(X_n) - g(c)| > \varepsilon\} \cap \{Y_n < \eta\}$, we have $|X_n - c| \leq Y_n < \eta$ and then $|g(X_n) - g(c)| < \varepsilon/2$. This is impossible. Next, since $Y_n \rightarrow_{\mathbb{P}} 0$, we have

$$\lim_{n \rightarrow +\infty} \sup \mathbb{P}(|g(X_n) - g(c)| > \varepsilon) \leq \lim_{n \rightarrow +\infty} \sup \mathbb{P}(Y_n \geq \eta) = 0.$$

The proof is complete.

MAIN PROPERTIES.

- (1) If $X_n = o(1)$ a.s., then $X_n = o_{\mathbb{P}}(1)$.
- (2) $o_{\mathbb{P}}(a_n) = a_n o_{\mathbb{P}}(1)$ and $a_n o_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$.
- (3) $o_{\mathbb{P}}(a_n) o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n b_n)$.
- (4) $o_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$.
- (5) If $b_n \geq a_n$ for all $n \geq 1$, $o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(b_n)$.
- (6) $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(a_n)$.

(7) $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n + b_n)$ and $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n \vee b_n)$ where $a_n \vee b_n = \max(a_n, b_n)$.

(8) If $X_n = O(1)$ a.s., then $X_n = O_{\mathbb{P}}(1)$.

(9) $O_{\mathbb{P}}(a_n) = a_n O_{\mathbb{P}}(1)$.

(10) $O_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n b_n)$.

(11) $O_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)$.

(12) If $b_n \geq a_n$ for all $n \geq 1$, $O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(b_n)$.

(13) $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n)$.

(14) $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n + b_n)$ and $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n \vee b_n)$ where $a_n \vee b_n = \max(a_n, b_n)$.

(15) $o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n b_n)$.

(16) $o_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$ and $O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$.

(17a) If $a_n = O_{\mathbb{P}}(b_n)$ that $o(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(b_n)$.

(17b) If $b_n = O_{\mathbb{P}}(a_n)$ that $o_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n)$.

(17c) If $b_n = o_{\mathbb{P}}(a_n)$, a.s., that $o(a_n) + O_{\mathbb{P}}(b_n) = o(b_n)$.

(18) $(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = o_{\mathbb{P}}(a_n)$.

(19) An $o_{\mathbb{P}}(1)$ is an $O_{\mathbb{P}}(1)$

PROOFS.

(1) This derived from the implication : $X_n \rightarrow 0$ a.s. $\implies X_n \rightarrow_P 0$ (See Proposition 11).

(2) If $X_n = o_{\mathbb{P}}(a_n) \iff |X_n/a_n| \rightarrow_P 0 \iff X_n/a_n = o_{\mathbb{P}}(1) \iff X_n = a_n o_{\mathbb{P}}(1)$.

(3) By Point (2) above, $o_{\mathbb{P}}(a_n)o_{\mathbb{P}}(b_n) = a_nb_n \times o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_nb_n \times o_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_nb_n)$ (Check $o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$ in Property (A1) in the Appendix subsection below).

(4) $o_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)o_{\mathbb{P}}(1) = a_n \times o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_n \times o_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$ (Use again Property (A1) the Appendix subsection below).

(5) Let $b_n \geq a_n$ for all $n \geq 1$, $X_n = o_{\mathbb{P}}(a_n)$. For any $\eta > 0$, $0 \leq \lim_{n \rightarrow +\infty} \sup P(|X_n/b_n| > \eta) \leq \lim_{n \rightarrow +\infty} P(|X_n/a_n| > \eta) = 0$.

(6) Let $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = o_{\mathbb{P}}(a_n)$. Use the classical stuff, for $\eta > 0$,

$$\left(\frac{|X_n|}{a_n} > \eta/2 \right) \cap \left(\frac{|Y_n|}{a_n} > \eta/2 \right) \subset \left(\frac{|X_n + Y_n|}{a_n} > \eta \right).$$

Then for $\eta > 0$,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n + Y_n|}{a_n} > \eta \right) &\leq \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n|}{a_n} > \eta/2 \right) \\ (2.9) \quad &+ \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|Y_n|}{a_n} > \eta/2 \right) = 0. \end{aligned}$$

(7) To prove this point, combine Points (5) and (6) above.

(8) $X_n = O(1)$ *a.s.* as $n \rightarrow +\infty$ means there exists Ω_0 measurable such that $\mathbb{P}(\Omega_0) = 1$ and for any $\omega \in \Omega_0$,

$$\limsup_{n \rightarrow +\infty} |X_n(\omega)| = \inf_{n \geq 1} \sup_{p \geq n} |X_p| = M(\omega) < +\infty.$$

We have for all $n \geq 1$,

$$\mathbb{P}(|X_n| > \lambda) \leq \mathbb{P} \left(\sup_{p \geq n} |X_p| > \lambda \right).$$

We have $Y_n = \sup_{p \geq n} |X_p| \downarrow M$ *a.s.* Then $Y_n 1_{\Omega_0} \rightarrow_{\mathbb{P}} M 1_{\Omega_0}$. We are dealing with real random variable and we may apply the weak convergence results. We get $Y_n 1_{\Omega_0} \rightarrow_w M 1_{\Omega_0}$ by Proposition 11. By Theorem 4, we have for any continuity point of $F_M(\lambda) = P(M 1_{\Omega_0} \leq \lambda)$. Use the Monotone Convergence Theorem to get

$$\begin{aligned}
\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) &= \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\sup_{p \geq n} |X_p| 1_{\Omega_0} > \lambda\right) \\
&= \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\sup_{p \geq n} |X_p| 1_{\Omega_0} > \lambda\right) \\
&= \mathbb{P}(M 1_{\Omega_0} > \lambda) \\
&= 1 - F_M(\lambda).
\end{aligned}$$

Since the set of discontinuity points of F_M is at most countable (see Point 6 of Chapter 4). We then apply the formula above for $\lambda \rightarrow +\infty$ while λ are continuity points. Since $1 - F_M(\lambda) \rightarrow 0$ as $\lambda \rightarrow +\infty$, then for any $\varepsilon > 0$, we are able to pick one value of $\lambda(\varepsilon)$ which is a continuity point of F_M satisfying $1 - F_M(\lambda) < \varepsilon$. For any $\varepsilon > 0$, we have found $\lambda(\varepsilon) > 0$ such that

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda(\varepsilon)) \leq \varepsilon,$$

which implies

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) = 0.$$

Then $X_n = O_{\mathbb{P}}(1)$.

(9) Let $X_n = O_{\mathbb{P}}(a_n)$. Then

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/a_n| > \lambda) = 0.$$

This is the definition that $X_n/a_n = O_{\mathbb{P}}(1)$ and then $X_n = a_n O_{\mathbb{P}}(1)$.

(10) $O_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = a_n b_n O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = a_n b_n O_{\mathbb{P}}(1)$ (Check that $O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = O_{\mathbb{P}}(1)$ in Property (A2) in the Appendix subsection).

(11) $O_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)O_{\mathbb{P}}(1) = a_n O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = O_{\mathbb{P}}(a_n)$.

(12) Let $b_n \geq a_n$ for all $n \geq 1$ and $X_n = O_{\mathbb{P}}(a_n)$. Then

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/b_n| > \lambda) \leq \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/a_n| > \lambda) = 0.$$

Then $X_n = O_{\mathbb{P}}(b_n)$.

(13) Let $X_n = O_{\mathbb{P}}(a_n)$ and $Y_n = O_{\mathbb{P}}(a_n)$. Use the same technique as in Formula 2.9 below to get

$$(2.10) \quad \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n + Y_n|}{a_n} > \lambda \right) \leq \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n|}{a_n} > \lambda/2 \right) + \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|Y_n|}{a_n} > \lambda/2 \right) = 0.$$

(14) Combine Points (12) and (13) to get this one.

(15) $o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = a_nb_no_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = a_nb_no_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_nb_n)$. (Check that $o_{\mathbb{P}}(1)O_{\mathbb{P}}(1)$ in Property (A3) in the Appendix subsection below).

(16) $o_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)o_{\mathbb{P}}(1) = a_nO_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_no_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$ and $O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$.

(17a) Let $a_n = O(b_n)$ and $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = O_{\mathbb{P}}(b_n)$. There exists $C > 0$ such that $a_n \leq Cb_n$ for any $n \geq 1$. Then $X_n = o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(Cb_n)$ by Point (5). But obviously $X_n = o_{\mathbb{P}}(Cb_n) = o_{\mathbb{P}}(b_n)$ and then $X_n = O_{\mathbb{P}}(b_n)$ by Point (19) below. Finally $X_n + Y_n = O_{\mathbb{P}}(b_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(b_n)$.

(17b) Let $b_n = O(a_n)$ and $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = O_{\mathbb{P}}(b_n)$. We exchange the roles of a_n and b_n to get $b_n \leq Ca_n$ and $Y_n = O_{\mathbb{P}}(Ca_n) = O_{\mathbb{P}}(a_n)$ by Point (1) and finally

$$X_n + Y_n = o_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n).$$

(17c) Let $b_n = o_{\mathbb{P}}(a_n)$ and $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = O_{\mathbb{P}}(b_n)$. We have $X_n + Y_n = o_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(a_n)$ by Point (16). Finally $X_n + Y_n = o_{\mathbb{P}}(a_n)$.

(18) We have

$$(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = \frac{o_{\mathbb{P}}(a_n)}{1 + o_{\mathbb{P}}(a_n)}.$$

By Point (b) of Lemma 8, $(1 + o_{\mathbb{P}}(a_n))^{-1} \rightarrow_P 1$ and by Point (a) of the same lemma, $(1 + o_{\mathbb{P}}(a_n))^{-1} = O_{\mathbb{P}}(1)$. Then

$$(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = O_{\mathbb{P}}(1)o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(a_n),$$

by Point (15).

(19) By Lemma 8, an $o_{\mathbb{P}}(1)$ converges to 0 in probability and then is an $O_{\mathbb{P}}(1)$.

2.1. Extensions. The concepts of *small o's* and *big O's* are extended to \mathbb{R}^k in the following way :

(1) The sequence of random vectors $(X_n)_{n \geq 1}$ in of \mathbb{R}^k , is an $o(a_n)$ *a.s.* if and only if $\|X_n\|/a_n = o(1)$ *a.s.*, and is an $o_{\mathbb{P}}(a_n)$ if and only if $\|X_n\|/a_n = o_{\mathbb{P}}(1)$.

(b) The sequence of random vectors $(X_n)_{n \geq 1}$ in of \mathbb{R}^k , is an $O(a_n)$ *a.s.* if and only if $\|X_n\|/a_n = O(1)$ *a.s.*, and is an $O_{\mathbb{P}}(a_n)$ if and only if $\|X_n\|/a_n = O_{\mathbb{P}}(1)$.

From there, handling these concepts is easy by combining their properties in \mathbb{R} and those of the norms in \mathbb{R}^k .

2.2. Balanced sequences. It may help in some cases to have sequences X_n such both $\|X_n\|$ and $1/\|X_n\|$ are bounded in probability. Let us give some notations for real sequences.

(1) For $0 \leq a < b < +\infty$, we denote by $X_n = O_{\mathbb{P}}(a, b, a_n, b_n)$ the property that for ant $\varepsilon > 0$, there exists $\lambda > 0$ such that we have $(a + \lambda \leq |X_n|/a_n, |X_n|/a_n \leq b - \lambda)$ *WPALE* $(1 - \varepsilon)$, for large values of n . If $a_n = b_n$ for all $n \geq 1$, we simply write $X_n = O_{\mathbb{P}}(a, b, a_n)$.

(1) For $0 \leq a$, we denote by $X_n = O_{\mathbb{P}}(a, +\infty, a_n, b_n)$ the property that for ant $\varepsilon > 0$, there exists $\lambda > 0$ such that we have

$$(a + \lambda^{-1} \leq |X_n|/a_n, |X_n|/a_n \leq \lambda)$$

WPALE $(1 - \varepsilon)$, for large values of n . If $a_n = b_n$ for all $n \geq 1$, we simply write $X_n = O_{\mathbb{P}}(a, +\infty, a_n)$.

An example of a sequence of random variables satisfying $X_n = O_{\mathbb{P}}(0, +\infty, 1)$ is a sequence X_n weakly converging to $X > 0$ *a.s.* In this case $1/X_n \rightsquigarrow$

$1/X$ finite *a.s.* and then $X_n = O_{\mathbb{P}}(1)$ and $1/X_n = O_{\mathbb{P}}(1)$. Combining these two points leads to $X_n = O_{\mathbb{P}}(0, +\infty, 1)$.

2.3. Appendix. (A1) If $X_n \rightarrow_{\mathbb{P}} a \in \mathbb{R}$ and $X_n \rightarrow_{\mathbb{P}} b \in \mathbb{R}$, then $X_n Y_n \rightarrow_{\mathbb{P}} ab$.

Proof. We have $(\eta + |b|)\eta + |a|\eta \rightarrow 0$ as $\eta \rightarrow 0$. For any $\varepsilon > 0$, for any $\delta > 0$, choose a value of $\eta > 0$ such that $(\eta + |b|)\eta + |a|\eta < \delta$. We apply the definition of the convergences $X_n \rightarrow_{\mathbb{P}} a$ and $X_n \rightarrow_{\mathbb{P}} b$ to get a value $N_0 \geq 1$ such that for any $n \geq N_0$,

$$\mathbb{P}(|X_n - a| \geq \eta) \leq \varepsilon/2 \text{ and } \mathbb{P}(|Y_n - b| \geq \eta) \leq \varepsilon/2.$$

But

$$\begin{aligned} |X_n Y_n - ab| &= |X_n Y_n - a Y_n + a Y_n - ab| \\ &\leq |Y_n| |X_n - a| + |a| |Y_n - b| \\ &\leq (|Y_n - b| + |b|) |X_n - a| + |a| |Y_n - b| \end{aligned}$$

On $(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c$,

$$|X_n Y_n - ab| \leq (\eta + |b|)\eta + |a|\eta \leq \delta.$$

Then for $n \geq N_0$,

$$(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c \subset (|X_n Y_n - ab| \leq \delta),$$

that is

$$\mathbb{P}(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c \leq \mathbb{P}(|X_n Y_n - ab| \leq \delta),$$

and by taking complementations,

$$\mathbb{P}(|X_n Y_n - ab| > \delta) \leq \mathbb{P}((|X_n - a| \geq \eta) \cup (|Y_n - b| \geq \eta)) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Then

$$X_n Y_n \rightarrow_{\mathbb{P}} ab.$$

Property (A2). $X_n = O_{\mathbb{P}}(1)$ and $X_n = O_{\mathbb{P}}(1)$ then $X_n Y_n = O_{\mathbb{P}}(1)$.

Proof. By applying the definition of an $O_{\mathbb{P}}(1)$, we may find for any $\varepsilon > 0$, two integers numbers N_1 and N_2 and two positive numbers $\lambda_1 > 0$ and $\lambda_2 > 0$ such that

$$\forall (n \geq N_1), \mathbb{P}(|X_n| \leq \lambda_1) \geq 1 - \varepsilon/2 \text{ and } \forall (n \geq N_2), \mathbb{P}(|Y_n| \leq \lambda_2) \geq 1 - \varepsilon/2.$$

For $n \geq \max(N_1, N_2)$,

$$(|X_n| \leq \lambda_1) \cap (|Y_n| \leq \lambda_2) \subset (|X_n Y_n| \leq \lambda_1 \lambda_2)$$

which is equivalent to

$$(|X_n Y_n| > \lambda_1 \lambda_2) \subset (|X_n| > \lambda_1) \cup (|Y_n| > \lambda_2)$$

which implies for $n \geq \max(N_1, N_2)$,

$$\mathbb{P}(|X_n Y_n| > \lambda_1 \lambda_2) \leq \mathbb{P}(|X_n| > \lambda_1) + \mathbb{P}(|Y_n| > \lambda_2) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Then for any $\varepsilon > 0$, there exists a nonnegative $N (= \max(N_1, N_2))$, there exists $\lambda > 0 (= \lambda_1 \lambda_2)$ and for any $n \geq N$,

$$\mathbb{P}(|X_n Y_n| \leq \lambda) \geq 1 - \varepsilon.$$

Hence $X_n Y_n = O_{\mathbb{P}}(1)$.

$$X_n Y_n \rightarrow_{\mathbb{P}} ab.$$

Property (A3). If $X_n = o_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$ then $X_n Y_n = o_{\mathbb{P}}(1)$.

Proof. Fix $\varepsilon > 0$. By applying the definition of an $O_{\mathbb{P}}(1)$ there exist an integer number N_1 and a positive number $\lambda > 0$ such that

$$\mathbb{P}(|Y_n| \leq \lambda) \geq 1 - \varepsilon/2.$$

Now let $\eta > 0$. Let us apply the definition of $X_n = o_{\mathbb{P}}(1)$ to get that there exists a positive integer N_2 such that

$$\forall (n \geq N_2), \mathbb{P}(|X_n| > \eta/\lambda) \leq \varepsilon/2.$$

Thus for $n \geq \max(N_1, N_2)$,

$$(|X_n| \leq \eta/\lambda) \cap (|Y_n| \leq \lambda) \subset (|X_n Y_n| \leq \eta)$$

which is equivalent to

$$(|X_n Y_n| > \eta) \subset (|X_n| > \eta/\lambda) \cup (|Y_n| > \lambda)$$

which implies for $n \geq \max(N_1, N_2)$,

$$\mathbb{P}(|X_n Y_n| > \eta) \leq \mathbb{P}(|X_n| > \eta/\lambda) + \mathbb{P}(|Y_n| > \lambda) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Then for any $\varepsilon > 0$, for any $\eta > 0$, there exists a nonnegative $N (= \max(N_1, N_2))$, for any $n \geq N$,

$$\mathbb{P}(|X_n Y_n| > \eta) \leq \varepsilon.$$

Hence $X_n Y_n = o_{\mathbb{P}}(1)$.

3. Delta Methods

The Delta method is a quick way to derive new asymptotic laws for sequences of random variables defined on the same probability measure $(\Omega, \mathcal{A}, \mathbb{P})$. Here, we present the univariate and multivariate case. We will see here the usefulness of the results in Section 6 of Chapter 2 combined with the manipulations of the o 's and the O 's in probability we just exposed in the first section of this chapter.

We begin by Delta Methods in \mathbb{R} .

3.1. Univariate Version.

PROPOSITION 32. *Let $(X_n)_{n \geq 1}$ be a sequence real random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let θ be a real number and $(a_n > 0)_{n \geq 1}$ be sequence of real numbers such that $a_n \rightarrow +\infty$ as $n \rightarrow +\infty$.*

Let $g : D \rightarrow \mathbb{R}$ be a function of class C^1 , such that D is a domain of \mathbb{R} , θ is in the interior $\overset{\circ}{D}$ of D , $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$.

If $a_n(X_n - \theta)$ weakly converges to a random variable Z as $n \rightarrow +\infty$, then

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow g'(\theta)Z \text{ as } n \rightarrow +\infty,$$

where $\nabla g(a) = g'(\theta)$ is the derivative of g at θ .

Proof of Proposition 32. Assume that all the hypotheses of the proposition are true. By Point (a) of Lemma 8, we have $a_n(X_n - \theta) = O_P(1)$ and then

$$X_n = \theta + O_P(1)a_n^{-1} \rightarrow_{\mathbb{P}} \theta$$

which by Proposition 13 in Section 6 of Chapter 2, is equivalent to the weak convergence

$$X_n \rightsquigarrow \theta.$$

Now the Mean Value Theorem implies

$$(3.1) \quad g(X_n) - g(\theta) = g'(Y_n)(X_n - \theta),$$

where

$$\min(X_n, \theta) \leq Y_n \leq \max(X_n, \theta),$$

that is

$$|Y_n - \theta| \leq |X_n - \theta|.$$

It comes that $Y_n \rightarrow_{\mathbb{P}} \theta$ and since g' is continuous, we have $g'(Y_n) \rightarrow_{\mathbb{P}} g'(\theta)$ by Point (b) of Lemma 8. Then by using Proposition 13 in Section 6 of Chapter 2, we see that is equivalent to

$$g'(Y_n) \rightsquigarrow g'(\theta).$$

By the property of Slutsky given in 15 in Section 6 of Chapter 2, we have the weak convergence

$$(g'(Y_n), a_n(X_n - \theta)) \rightsquigarrow (g'(\theta), Z)$$

and by the continuous mapping Theorem 6 in Chapter 2 combined with (3.1), we get the final conclusion

$$a_n(g(X_n) - g(\theta)) = (g'(Y_n) \times a_n(X_n - \theta)) \rightsquigarrow g'(\theta)Z.$$

Remark. Let us use the derivative map (total derivative)

$$h \rightarrow g'_\theta(h) = g'(\theta)h,$$

in Proposition 32, we write the conclusion in the form

$$a_n(g(X_n) - g(\theta)) = g'_\theta(Z).$$

This writing suggests we may have this kind of results in more general spaces. Let us move to the multivariate case.

3.2. Multivariate version. The first statement concerns the transformation of the converging sequence of k components by a real function of k arguments.

PROPOSITION 33. *Let $(X_n)_{n \geq 1}$ be a sequence k -random vectors, $k \geq 1$, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\theta \in \mathbb{R}^k$ and $(a_n > 0)_{n \geq 1}$ be sequence of real numbers such that $a_n \rightarrow +\infty$ as $n \rightarrow +\infty$.*

Let $g : D \rightarrow \mathbb{R}$ be a function of class C^1 , such that D is a domain of \mathbb{R}^k , θ is $\overset{\circ}{D}$, the interior of D , $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$.

If $a_n(X_n - \theta)$ weakly converges to a k -random vector Z as $n \rightarrow +\infty$, then

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow {}^t\nabla g(\theta)Z = \langle \nabla g(\theta), Z \rangle \text{ as } n \rightarrow +\infty,$$

where

$${}^t\nabla g(\theta) = \left(\frac{\partial g(\theta)}{\partial \theta_1}, \dots, \frac{\partial g(\theta)}{\partial \theta_k} \right)$$

is the gradient vector of g at θ .

The second statement is the most general in the finite dimension frame, in which the converging sequence of k components is transformed by a multicomponent function of k arguments.

PROPOSITION 34. *Let $(X_n)_{n \geq 1}$ be a sequence k -random vectors, $k \geq 1$, defined on the same probability space (Ω, A, \mathbb{P}) and let θ be a real number and $(a_n > 0)_{n \geq 1}$ be sequence of real numbers such that $a_n \rightarrow +\infty$ as $n \rightarrow +\infty$.*

Let $g : D \rightarrow \mathbb{R}^m$ be function of class C^1 , such that θ is $\overset{\circ}{D}$, the interior $\overset{\circ}{D}$ of D , $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$. Denote by g_j , $1 \leq j \leq m$, the components of the function g .

If $a_n(X_n - \theta)$ weakly converges to a k -random vector Z as $n \rightarrow +\infty$, then

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow g'_\theta Z = \text{ as } n \rightarrow +\infty,$$

where g'_θ is the matrix of partial derivatives of first order

$$g'_\theta = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_j} & \cdots & \frac{\partial g_1}{\partial \theta_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_i}{\partial \theta_1} & \cdots & \frac{\partial g_i}{\partial \theta_j} & \cdots & \frac{\partial g_i}{\partial \theta_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_m}{\partial \theta_1} & \cdots & \frac{\partial g_m}{\partial \theta_j} & \cdots & \frac{\partial g_m}{\partial \theta_k} \end{pmatrix},$$

Proof of Proposition 33. Assume that the hypotheses of the proposition hold.

Let us use the expansion of g of first order at $\theta = {}^t(\theta_1, \dots, \theta_k)$ for $x = {}^t(x_1, \dots, x_k)$

$$(3.2) \quad g(x) - g(\theta) = (x_1 - \theta_1) \frac{\partial g}{\partial \theta_1}(\theta) + \dots + (x_k - \theta_k) \frac{\partial g}{\partial \theta_k}(\theta) + o(\|x - \theta\|).$$

Since $a_n(T_n - \theta) = a_n^t((T_{1,n}, \dots, T_{k,n}) - {}^t(\theta_1, \dots, \theta_k)) \rightsquigarrow Z = {}^t(Z_1, \dots, Z_k)$, it follows from the continuous mapping theorem 6 of Chapter 2, that for each $1 \leq i \leq k$, $a_n(T_{i,n} - \theta_i)$ converges to Z_i and then Point (a) of lemme 8, we get

$$1 \leq j \leq k, (T_{j,n} - \theta_j) = O_{\mathbb{P}}(a_n^{-1}).$$

Then by Points (10) and (13) of the main properties in Part II of the above section

$$(3.3) \quad \|T_n - \theta\| = \left\{ \sum_{j=1}^k (T_{j,n} - \theta_j)^2 \right\}^{1/2} = O_{\mathbb{P}}(a_n^{-1}) = o_{\mathbb{P}}(1).$$

The term $o(\|x - \theta\|)$ in 3.2 is continuous as a difference of two continuous functions and takes the value 0 for $\|x - \theta\| = 0$. By using Part (c) of Lemma 8, a combination of (3.2) and (3.3) leads to

$$\begin{aligned} a_n(g(x) - g(\theta)) &= a_n(T_{1,n} - \theta_1) \frac{\partial g}{\partial \theta_1}(\theta) + \dots + a_n(T_{k,n} - \theta_k) \frac{\partial g}{\partial \theta_k}(\theta) + a_n o_{\mathbb{P}}(O_{\mathbb{P}}(a_n^{-1})) \\ &= {}^t \nabla g(\theta)(a_n(T_n - \theta) + o_{\mathbb{P}}(1)). \end{aligned}$$

This says that $a_n(g(x) - g(\theta))$ and ${}^t \nabla g(\theta)(a_n(T_n - \theta))$ are equivalent in probability. Since ${}^t \nabla g(\theta)(a_n(T_n - \theta)) \rightsquigarrow {}^t \nabla g(\theta)Z$ by the continuous mapping Theorem, we get by Proposition 14 in Section 6 of Chapter 2,

$$a_n(g(x) - g(\theta)) \rightsquigarrow {}^t \nabla g(\theta)Z.$$

Proof of Proposition 34. Assume that the hypotheses of the proposition hold.

The function g has m components $g_j \in \mathbb{R}^m$ so that we write $g = {}^t(g_1, \dots, g_m)$. Each component is of class C^1 . Let us use the conclusion of Proposition 33 for each of these components at $\theta = {}^t(\theta_1, \dots, \theta_k)$ for $x = {}^t(x_1, \dots, x_k)$ to get

$$(3.4) \quad g_j(x) - g_j(\theta) = (x_1 - \theta_1) \frac{\partial g_j}{\partial \theta_1}(\theta) + \dots + (x_k - \theta_k) \frac{\partial g_j}{\partial \theta_k}(\theta) + o(\|x - \theta\|).$$

This can be written using matrices as

$$g(x) - g(\theta) = g'_{\theta}(x - \theta) + o^{(m)}(\|x - \theta\|),$$

where $o^{(m)}(\|x - \theta\|)$ is a vector of m coordinates such that each of them is a continuous function which is also an $o(\|x - \theta\|)$. A similar notation is also used the $o_{\mathbb{P}}(\circ)$. By applying the method used in Proposition 33, we get

$$g(T_n) - g(\theta) = g'_{\theta}(T_n - \theta) + o_{\mathbb{P}}^{(m)}(a_n^{-1})$$

and

$$a_n(g(T_n) - g(\theta)) = g'_{\theta} a_n(T_n - \theta) + o_{\mathbb{P}}^{(m)}(1).$$

We have

$$\|a_n(g(T_n) - g(\theta)) - g'_\theta a_n(T_n - \theta)\|_{\mathbb{R}^m} = \left\| o_{\mathbb{P}}^{(m)}(1) \right\|_{\mathbb{R}^m} = o_{\mathbb{P}}(1).$$

Then $a_n(g(T_n) - g(\theta))$ has the same weak limit as $g'_\theta a_n(T_n - \theta)$ which is $g'_\theta Z$ by the continuous mapping.

4. Using the Functional Empirical Process in Asymptotic Statistics

4.1. The Functional empirical process. The functional empirical process (FEP) is a powerful tool for deriving asymptotic limit distributions. It is similar to the multivariate delta method. But the PEF has an advantage we describe below.

Given a sequence Z_1, Z_2, \dots , if independent and identically distributed random variables, of common probability law \mathbb{P}_0 , we will be able

(1) to find one Gaussian stochastic process $\mathbb{G}_{\mathbb{P}_0}$

and

(2) to express the asymptotic distributions of statistics which functions of Z_1, Z_2, \dots, Z_n with respect to $\mathbb{G}_{\mathbb{P}_0}$.

This allows to study separately all statistics based on Z_1, Z_2, \dots, Z_n and, each time we want it, to get the joint asymptotic distributions of any finite number of them. **We say that we place the asymptotic distributions of statistics based on Z_1, Z_2, \dots, Z_n in the Gaussian field of $\mathbb{G}_{\mathbb{P}_0}$.**

An other interesting point is that the joint distributions we obtain by using the *FEP* tool, have their covariance functions expressed in functional forms. Whatever be complicated these covariances, we do not have to worry about their form since the powerful computers of modern times are able to compute them in very short times.

The Delat method does not have this unified frame. Instead, each work is done for once. When we need to add or drop one statistics, we have to do the job again.

Before we present the functional empirical process, we want to reassure the reader that we will only use finite distributions of the functional empirical process, that is, we remain in \mathbb{R}^k and we will not use the heavy tools of functional topologies or Vapnick-Cervonenkis classes.

Let Z_1, Z_2, \dots be a sequence of independent copies of a random variable Z defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values on some

metric space (S, d) . Define for each $n \geq 1$, the functional empirical process by

$$\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(Z_i) - \mathbb{E}f(Z_i)),$$

where f is a real and measurable function defined on \mathbb{R} such that

$$(4.1) \quad \mathbb{V}_Z(f) = \int (f(x) - \mathbb{P}_Z(f))^2 dP_Z(x) < \infty,$$

which entails

$$(4.2) \quad \mathbb{P}_Z(|f|) = \int |f(x)| dP_Z(x) < \infty.$$

Denote by $\mathcal{F}(S)$ - \mathcal{F} for short - the class of real-valued measurable functions that are defined on S such that (4.1) holds. The space \mathcal{F} , when endowed with the addition and the external multiplication by real scalars, is a linear space. Next, it remarkabke that \mathbb{G}_n is linear on \mathcal{F} , that is for f and g in \mathcal{F} and for $(a, b) \in \mathbb{R}^2$, we have

$$a\mathbb{G}_n(f) + b\mathbb{G}_n(g) = \mathbb{G}_n(af + bg).$$

We have this result

LEMMA 10. *Given the notation above, then for any finite number of elements f_1, \dots, f_k of \mathcal{S} , $k \geq 1$, we have*

$${}^t(\mathbb{G}_n(f_1), \dots, \mathbb{G}_n(f_k)) \rightsquigarrow \mathcal{N}_k(0)$$

where

$$\Gamma(f_i, f_j) = \int (f_i - \mathbb{P}_Z(f_i))(f_j - \mathbb{P}_Z(f_j)) dP_Z(x), 1 \leq, j \leq k.$$

This lemma says that the weak limit of the sequence ${}^t(\mathbb{G}_n(f_1), \mathbb{G}_n(f_2), \dots, \mathbb{G}_n(f_k))$ has the same law than the vector ${}^t(\mathbb{G}(f_1), \mathbb{G}(f_2), \dots, \mathbb{G}(f_k))$, where $\{\mathbb{G}(f), f \in \mathcal{F}\}$ is a Gaussian process of variance-covariance function

$$(4.3) \quad \Gamma(f, g) = \int (f - \mathbb{P}_Z(f))(g - \mathbb{P}_Z(g)) dP_Z(x), \quad (f, g) \in \mathcal{F}^2.$$

By applying the Skorohod-Wichura Theorem (See Chapter 2), we may suppose that we are on the same probability space on which we have the following approximation :

$$(4.4) \quad \mathbb{G}_n(f_1) = \mathbb{G}_n(f_1) + o_{\mathbb{P}}(1), 1 \leq i \leq p.$$

We will come back later on the application of the formula.

PROOF. It is enough to use the Cramér-Wold Criterion (see Proposition 1 in Chapter 1), that is to show that for any $a = {}^t(a_1, \dots, a_k) \in \mathbb{R}^k$, we have

$$\langle a, T_n \rangle \rightsquigarrow \langle a, T \rangle$$

where we have used the notation $T_n = {}^t(\mathbb{G}_n(f_1), \dots, \mathbb{G}_n(f_k))$, and where T follows the $\mathcal{N}_k(0, \Gamma(f_i, f_j)_{1 \leq i, j \leq k})$ law, and $\langle \circ, \circ \rangle$ stands for the usual product scalar in \mathbb{R}^k .

But, by the standard central limit theorem in \mathbb{R} , we have

$$\langle a, T_n \rangle = \mathbb{G}_n \left(\sum_{i=1}^k a_i f_i \right) \rightsquigarrow \mathcal{N}(0, \sigma_{\infty}^2),$$

where, for $g = \sum_{1 \leq i \leq k} a_i f_i$,

$$\sigma_{\infty}^2 = \int (g(x) - \mathbb{P}_Z(g))^2 dP_Z(x)$$

and this easily gives

$$\sigma_{\infty}^2 = \sum_{1 \leq i, j \leq k} a_i a_j \Gamma(f_i, f_j),$$

so that $\mathcal{N}(0, \sigma_{\infty}^2)$ is the law of $\langle a, T \rangle$. The proof is finished.

4.2. How to use the FEP tool? The usual statistics we are working with in Asymptotic Statistics are based on univariate or multivariate samples, meaning we usually work on \mathbb{R}^k . Once we have our sample Z_1, Z_2, \dots as random variables defined in the same probability space with values in \mathbb{R}^k , the studied statistic, say T_n , is usually a combination of expressions of the form

$$H_n = \frac{1}{n} \sum_{i=1}^k H(Z_i)$$

for $H \in \mathcal{F}$. We use the results of Lemma 10 and Point (a) of Lemma 8, to have this very sample expansion $\mu(H) = \mathbb{E}H(Z)$,

$$(4.5) \quad H_n = \mu(H) + n^{-1/2}\mathbb{G}_n(H).$$

We have that $\mathbb{G}_n(H)$ is asymptotically bounded in probability since $\mathbb{G}_n(H)$ weakly converges to, say $M(H)$ and then by the continuous mapping theorem $\|\mathbb{G}_n(H)\| \rightsquigarrow \|M(H)\|$. Since all the $\mathbb{G}_n(H)$ are defined on the same probability space, we get for all $\lambda > 0$, by the assertion of the Portmanteau Theorem for concerning open sets,

$$\limsup_{n \rightarrow \infty} P(\|\mathbb{G}_n(H)\| > \lambda) \leq P(\|M(H)\| > \lambda)$$

and then

$$\liminf_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\|\mathbb{G}_n(H)\| > \lambda) \leq \liminf_{\lambda \rightarrow \infty} \mathbb{P}(\|M(H)\| > \lambda) = 0.$$

From this, we use the big $O_{\mathbb{P}}$ notation, that is $\mathbb{G}_n(H) = O_{\mathbb{P}}(1)$. Formula (4.5) becomes

$$H_n = \mu(H) + n^{-1/2}\mathbb{G}_n(H) = \mu(H) + O_{\mathbb{P}}(n^{-1/2})$$

and we will be able to use the delta method. Indeed, let $g : \mathbb{R} \rightarrow \mathbb{R}$ be continuously differentiable on a neighborhood of $\mu(H)$. The mean value theorem leads to

$$(4.6) \quad g(H_n) = g(\mu(H)) + g'(\mu_n(H)) n^{-1/2}\mathbb{G}_n(H)$$

where

$$\mu_n(H) \in [(\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \wedge \mu(H), (\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \vee \mu(H)]$$

so that

$$|\mu_n(H) - \mu(H)| \leq n^{-1/2}\mathbb{G}_n(H) = O_{\mathbb{P}}(n^{-1/2}).$$

Then $\mu_n(H)$ converges to $\mu(H)$ in probability (denoted $\mu_n(H) \rightarrow_{\mathbb{P}} \mu(H)$). But the convergence in probability to a constant is equivalent to the weak convergence. Then $\mu_n(H) \rightsquigarrow \mu(H)$. Using again the continuous mapping theorem, $g'(\mu_n(H)) \rightsquigarrow g'(\mu(H))$ which in turn yields $g'(\mu_n(H)) \rightarrow_{\mathbb{P}} g'(\mu(H))$ by the characterization of the weak convergence to a constant. Now (4.6) becomes

$$\begin{aligned}
g(H_n) &= g(\mu(H)) + (g'(\mu(H) + o_P(1)) n^{-1/2} \mathbb{G}_n(H) \\
&= g(\mu(H)) + g'(\mu(H) \times n^{-1/2} \mathbb{G}_n(H) + o_P(1)) n^{-1/2} \mathbb{G}_n(H) \\
&= g(\mu(H)) + n^{-1/2} \mathbb{G}_n(g'(\mu(H)H) + o_P(n^{-1/2}))
\end{aligned}$$

We arrive at the final expansion

$$(4.7) \quad g(H_n) = g(\mu(H)) + n^{-1/2} \mathbb{G}_n(g'(\mu(H)H) + o_P(n^{-1/2})).$$

By using the Skorohod-Wichura representation, we get by Formula , that

$$(4.8) \quad g(H_n) = g(\mu(H)) + n^{-1/2} \mathbb{G}(g'(\mu(H)H) + o_P(n^{-1/2})).$$

The method consists in using the expansion (4.7) as many times as needed and next to do some algebra on these expansions.

The algebraic computations we refereed above are based on the application of the following lemma.

LEMMA 11. *Let (A_n) and (B_n) be two sequences of real valued random variables defined on the same probability space holding the sequence Z_1, Z_2, \dots*

Let A and B be two real numbers and Let $L(z)$ and $H(z)$ be two real-valued functions of $z \in S$, with $(L, H) \in \mathcal{F}^2$.

Suppose that

$$A_n = A + n^{-1/2} \mathbb{G}_n(L) + o_P(n^{-1/2})$$

and

$$A_n = B + n^{-1/2} \mathbb{G}_n(H) + o_P(n^{-1/2}).$$

Then

$$A_n + B_n = A + B + n^{-1/2} \mathbb{G}_n(L + H) + o_P(n^{-1/2}),$$

and

$$A_n B_n = AB + n^{-1/2} \mathbb{G}_n(BL + AH)$$

and if $B \neq 0$,

$$\frac{A_n}{B_n} = \frac{A}{B} + n^{-1/2} \mathbb{G}_n\left(\frac{1}{B}L - \frac{A}{B^2}H\right) + o_P(n^{-1/2})$$

By putting together all the described steps in a smart way, the methodology will lead us to a final result of the form

$$T_n = t + n^{-1/2}\mathbb{G}_n(h) + o_P(n^{-1/2}),$$

which entails the weak convergence

$$\begin{aligned}\sqrt{n}(T_n - t) &= \mathbb{G}_n(h) + o_P(1) \rightsquigarrow \mathcal{N}(0, \Gamma(h, h)) \\ &= \mathbb{G}(h) + o_P(1).\end{aligned}$$

Now, we are going to show how to apply the methodology on the empirical linear correlation coefficient.

4.3. An Example. We are going to illustrate our tool on the plug-in estimator of the linear correlation coefficient of two random variable (X, Y) , with neither of X nor Y is degenerated, defined as follows

$$\rho = \frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2}$$

where

$$\mu_x = \int x dP_X(x), \mu_y = \int y dP_Y(y), \sigma_{xy} = \int (x - \mu_x)(y - \mu_y) dP_{(X,Y)}(x, y)$$

and

$$\sigma_x^2 = \int (x - \mu_x)^2 dP_X(x), \sigma_y^2 = \int (y - \mu_y)^2 dP_Y(y).$$

We also dismiss the case where $|\rho| = 1$, for which one of X and Y is an affine function of the other, meaning for example that we have $X = aY + b$ for some $(a, b) \in \mathbb{R}^2$.

It is clear that centering the variables X and Y at their expectations and normalizing them by their standard deviations σ_x and σ_y do not change the correlation coefficient ρ . So we may and do center X and Y at their expectations and normalize them so that we can and do assume that

$$\mu_x = \mu_y = 0, \sigma_x = \sigma_y = 1.$$

However, we will let these coefficient appear with their names and we only use their particular values at the conclusion stage.

Let us construct the plug-in estimator of ρ . To this end, let $(X_1, Y_1), (X_2, Y_2), \dots$ be a sequence independent observations of (X, Y) . For each $n \geq 1$, the plug-in estimator is the following

$$\rho_n = \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right\} \left\{ \frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X})^2 \times \sum_{i=1}^n (Y_i - \bar{Y})^2 \right\}^{-1/2}.$$

We are going to give the asymptotic theory of ρ_n as an estimator of ρ . Introduce the notation

$$\mu_{(p,x),(q,y)} = E((X - \mu_x)^p (Y - \mu_y)^q), \mu_{4,x} = E(X - \mu_x)^4, \mu_{4,y} = E(Y - \mu_y)^4$$

Here is the outcome of the application of the method.

THEOREM 13. *Suppose that neither of X and Y is degenerated and both have finite fourth moments and that X^3Y and XY^3 have finite expectations. Then, as $n \rightarrow \infty$,*

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2),$$

where

$$\begin{aligned} \sigma^2 = & \sigma_x^{-2} \sigma_y^{-2} (1 + \rho^2/2) \mu_{(2,x),(2,y)} + \rho^2 (\sigma_x^{-4} \mu_{4,x} + \sigma_y^{-4} \mu_{4,y})/4 \\ & - \rho (\sigma_x^{-3} \sigma_y^{-1} \mu_{(3,x),(1,y)} + \sigma_x^{-1} \sigma_y^{-3} \mu_{(1,x),(3,y)}) \end{aligned}$$

This result enables to test independence between X and Y , or to test non linear correlation in the following sense.

THEOREM 14. *Suppose that the assumptions of Theorem 13 hold. Then*

(1) *If X and Y are not linearly correlated, that is $\rho \neq 0$, we have*

$$\sqrt{n}\rho_n \rightsquigarrow N(0, \sigma_1^2),$$

where

$$\sigma_1^2 = \sigma_x^{-2} \sigma_y^{-2} \mu_{(2,x),(2,y)}.$$

(2) *If X and Y are independent, then $\rho = 0$, and*

$$\sqrt{n}\rho_n \rightsquigarrow N(0, 1)$$

Proofs. We are going to use the function empirical process based on the observations $(X_i, Y_i), i = 1, 2, \dots$ that are independent copies of (X, Y) . Write

$$\rho_n^2 = \frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y}}{\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2}} = \frac{A_n}{B_n}.$$

Let us say for once that all the functions of $Z = (X, Y)$ that will appear below are measurable and have finite second moments. Let us handle separately the numerator and denominator. To treat A_n , using the empirical process implies that

$$(4.9) \quad \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i Y_i = \mu_{xy} + n^{-1/2} G_n(p), \\ \bar{X} = \mu_x + n^{-1/2} G_n(\pi_1), \\ \bar{Y} = \mu_y + n^{-1/2} G_n(\pi_2), \end{cases}$$

where $p(x, y) = xy$, $\pi_1(x, y) = x$, $\pi_2(x, y) = y$. From there we use the fact that $G_n(g) = O_P(1)$ for $\mathbb{E}(g(X, Y)^2) < +\infty$ and get

$$(4.10) \quad A_n = \mu_{xy} + n^{-1/2} G_n(p) - (\mu_x + n^{-1/2} G_n(\pi_1))(\mu_y + n^{-1/2} G_n(\pi_2)).$$

This leads to

$$A_n = \sigma_{xy} + n^{-1/2} G_n(H_1) + o_P(n^{-1/2})$$

with

$$H_1(x, y) = p(x, y) - \mu_x \pi_2 - \mu_y \pi_1.$$

Next, we have to handle B_n . Since the roles of $\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2}$ and of $\left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2}$ are symmetrical, we treat one of them and extend the results to the other. Let us handle $\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2}$. The combination of (4.9) and the Delta method enables to get

$$\bar{X}^2 = (\mu_x + n^{-1/2} G_n(\pi_1))^2 = \mu_x^2 + 2\mu_x n^{-1/2} G_n(\pi_1) + o_P(n^{-1/2})$$

that is

$$\bar{X}^2 = (\mu_x + n^{-1/2} G_n(\pi_1))^2 = \mu_x^2 + n^{-1/2} G_n(2\mu_x \pi_1) + o_P(n^{-1/2}).$$

From there, we get

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 &= m_{2,x} + n^{-1/2} G_n(\pi_1^2) - \bar{X}^2 \\
&= m_{2,x} - \mu_x^2 + n^{-1/2} G_n(\pi_1^2 - 2\mu_x \pi_1) + o_P(n^{-1/2}) \\
&= \sigma_x^2 + n^{-1/2} G_n(\pi_1^2 - 2\mu_x \pi_1) + o_P(n^{-1/2}).
\end{aligned}$$

Using the Delta-method once again leads to

$$\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} = \sigma_x + n^{-1/2} G_n\left(\frac{1}{2\sigma_x} \{\pi_1^2 - 2\mu_x \pi_1\}\right) + o_P(n^{-1/2}).$$

In a similar way, we get

$$\left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2} = \sigma_y + n^{-1/2} G_n\left(\frac{1}{2\sigma_y} \{\pi_2^2 - 2\mu_y \pi_2\}\right) + o_P(n^{-1/2}).$$

We arrive at

$$\begin{aligned}
B_n &= \left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2} \\
&= \sigma_x \sigma_y + n^{-1/2} G_n\left(\frac{\sigma_y}{2\sigma_x} \{\pi_1^2 - 2\mu_x \pi_1\} + \frac{\sigma_x}{2\sigma_y} \{\pi_2^2 - 2\mu_y \pi_2\}\right) + o_P(n^{-1/2}).
\end{aligned}$$

By setting

$$H_2(x, y) = \frac{\sigma_y}{2\sigma_x} \{\pi_1^2 - 2\mu_x \pi_1\} + \frac{\sigma_x}{2\sigma_y} \{\pi_2^2 - 2\mu_y \pi_2\}$$

we have

$$(4.11) \quad B_n = \sigma_x \sigma_y + n^{-1/2} G_n(H_2) + n^{-1/2}.$$

Now, combining (4.10) and (4.11) and using Lemma 11 yields

$$\sqrt{n}(\rho_n^2 - \rho^2) = n^{-1/2} G_n\left(\frac{1}{\sigma_x \sigma_y} H_1 - \frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2} H_2\right) + o_P(1).$$

Put

$$H = \frac{1}{\sigma_x \sigma_y} (p(x, y) - \mu_x \pi_2 - \mu_y \pi_1) - \frac{\rho}{\sigma_x \sigma_y} \left\{ \frac{1}{2\sigma_x^2} \{\pi_1^2 - 2\mu_x \pi_1\} + \frac{1}{2\sigma_y^2} \{\pi_2^2 - 2\mu_y \pi_2\} \right\}.$$

Now we continue with the centered and normalized case to get

$$H(x, y) = p(x, y) - \frac{\rho}{2}(\pi_1^2 + \pi_2^2)$$

and

$$H(X, Y) = XY - \frac{\rho}{2}(X^2 + Y^2).$$

Denote

$$\mu_{(p,x),(q,y)} = E((X - \mu_x)^p(Y - \mu_y)^q).$$

We have

$$\mathbb{E}H(X, Y) = \sigma_{xy} - \rho = 0$$

and $\text{var}H(X, Y)$ is equal to

$$\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}) + \rho^2\mu_{(2,x),(2,y)}/2$$

and finally

$$\text{var}H(X, Y) = \sigma_0^2$$

with

$$\sigma_0^2 = (1 + \rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}).$$

This gives the conclusion that for centered and normalized X and Y ,

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma_0^2).$$

Next, if we use the normalizing coefficients in σ_0 , we get

$$\begin{aligned} \sigma^2 &= \sigma_x^2\sigma_y^2(1 + \rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\sigma_x^4\mu_{4,x} + \sigma_y^4\mu_{4,y})/4 \\ &\quad - \rho(\sigma_x^3\sigma_y\mu_{(3,x),(1,y)} + \sigma_x\sigma_y^3\mu_{(1,x),(3,y)}) \end{aligned}$$

and we conclude in the general case that

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2)$$

The proof of Theorem 14 follows by easy computations under the particular conditions of ρ and under independence.

Elements of Theory of Functions and Real Analysis

1. Review on limits in $\overline{\mathbb{R}}$. What should not be ignored on limits.

Definition $\ell \in \overline{\mathbb{R}}$ is an accumulation point of a sequence $(x_n)_{n \geq 0}$ of real numbers finite or infinite, in $\overline{\mathbb{R}}$, if and only if there exists a subsequence $(x_{n(k)})_{k \geq 0}$ of $(x_n)_{n \geq 0}$ such that $x_{n(k)}$ converges to ℓ , as $k \rightarrow +\infty$.

Exercise 1 : Set $y_n = \inf_{p \geq n} x_p$ and $z_n = \sup_{p \geq n} x_p$ for all $n \geq 0$. Show that :

(1) $\forall n \geq 0, y_n \leq x_n \leq z_n$

(2) Justify the existence of the limit of y_n called limit inferior of the sequence $(x_n)_{n \geq 0}$, denoted by $\liminf x_n$ or $\underline{\lim} x_n$, and that it is equal to the following

$$\underline{\lim} x_n = \liminf x_n = \sup_{n \geq 0} \inf_{p \geq n} x_p$$

(3) Justify the existence of the limit of z_n called limit superior of the sequence $(x_n)_{n \geq 0}$ denoted by $\limsup x_n$ or $\overline{\lim} x_n$, and that it is equal

$$\overline{\lim} x_n = \limsup x_n = \inf_{n \geq 0} \sup_{p \geq n} x_p$$

(4) Establish that

$$-\liminf x_n = \limsup(-x_n) \quad \text{and} \quad -\limsup x_n = \liminf(-x_n).$$

(5) Show that the limit superior is sub-additive and the limit inferior is super-additive, i.e. : for two sequences $(s_n)_{n \geq 0}$ and $(t_n)_{n \geq 0}$

$$\limsup(s_n + t_n) \leq \limsup s_n + \limsup t_n$$

and

$$\liminf(s_n + t_n) \geq \liminf s_n + \liminf t_n$$

(6) Deduce from (1) that if

$$\liminf x_n = \limsup x_n,$$

then $(x_n)_{n \geq 0}$ has a limit and

$$\lim x_n = \liminf x_n = \limsup x_n$$

Exercise 2. Accumulation points of $(x_n)_{n \geq 0}$.

(a) Show that if $\ell_1 = \liminf x_n$ and $\ell_2 = \limsup x_n$ are accumulation points of $(x_n)_{n \geq 0}$. Show one case and deduce the second using point (3) of exercise 1.

(b) Show that ℓ_1 is the smallest accumulation point of $(x_n)_{n \geq 0}$ and ℓ_2 is the biggest. (Similarly, show one case and deduce the second using point (3) of exercise 1).

(c) Deduce from (a) that if $(x_n)_{n \geq 0}$ has a limit ℓ , then it is equal to the unique accumulation point and so,

$$\ell = \overline{\lim} x_n = \limsup x_n = \inf_{n \geq 0} \sup_{p \geq n} x_p.$$

(d) Combine this result with point (6) of Exercise 1 to show that a sequence $(x_n)_{n \geq 0}$ of \mathbb{R} has a limit ℓ in \mathbb{R} if and only if $\liminf x_n = \limsup x_n$ and then

$$\ell = \lim x_n = \liminf x_n = \limsup x_n$$

Exercise 3. Let $(x_n)_{n \geq 0}$ be a non-decreasing sequence of $\overline{\mathbb{R}}$. Study its limit superior and its limit inferior and deduce that

$$\lim x_n = \sup_{n \geq 0} x_n.$$

Deduce that for a non-increasing sequence $(x_n)_{n \geq 0}$ of $\overline{\mathbb{R}}$,

$$\lim x_n = \inf_{n \geq 0} x_n.$$

Point 4. (Convergence criteria)

Prohorov Criterion Let $(x_n)_{n \geq 0}$ be a sequence of $\overline{\mathbb{R}}$ and a real number $\ell \in \overline{\mathbb{R}}$ such that: Every subsequence of $(x_n)_{n \geq 0}$ also has a subsequence (that is a subsubsequence of $(x_n)_{n \geq 0}$) that converges to ℓ . Then, the limit of $(x_n)_{n \geq 0}$ exists and is equal ℓ .

Upcrossing or Downcrossing Criterion. Upcrossings and downcrossings.

Let $(x_n)_{n \geq 0}$ be a sequence in $\overline{\mathbb{R}}$ and two real numbers a and b such that $a < b$. We define

$$\nu_1 = \begin{cases} \inf & \{n \geq 0, x_n < a\} \\ +\infty & \text{if } (\forall n \geq 0, x_n \geq a) \end{cases}.$$

If ν_1 is finite, let

$$\nu_2 = \begin{cases} \inf & \{n > \nu_1, x_n > b\} \\ +\infty & \text{if } (n > \nu_1, x_n \leq b) \end{cases}.$$

As long as the ν'_j s are finite, we can define for ν_{2k-2} ($k \geq 2$)

$$\nu_{2k-1} = \begin{cases} \inf & \{n > \nu_{2k-2}, x_n < a\} \\ +\infty & \text{if } (\forall n > \nu_{2k-2}, x_n \geq a) \end{cases}.$$

and for ν_{2k-1} finite,

$$\nu_{2k} = \begin{cases} \inf & \{n > \nu_{2k-1}, x_n > b\} \\ +\infty & \text{if } (n > \nu_{2k-1}, x_n \leq b) \end{cases}.$$

We stop once one ν_j is $+\infty$. If ν_{2j} is finite, then

$$x_{\nu_{2j}} - x_{\nu_{2j-1}} > b - a.$$

We then say : by that moving from $x_{\nu_{2j-1}}$ to $x_{\nu_{2j}}$, we have accomplished a crossing (toward the up) of the segment $[a, b]$ called *up-crossings*. Similarly, if one ν_{2j+1} is finite, then the segment $[x_{\nu_{2j}}, x_{\nu_{2j+1}}]$ is a crossing

downward (downcrossing) of the segment $[a, b]$. Let

$D(a, b) =$ number of upcrossings of the sequence of the segment $[a, b]$.

(a) What is the value of $D(a, b)$ if ν_{2k} is finite and ν_{2k+1} infinite.

(b) What is the value of $D(a, b)$ if ν_{2k+1} is finite and ν_{2k+2} infinite.

(c) What is the value of $D(a, b)$ if all the ν'_j 's are finite.

(d) Show that $(x_n)_{n \geq 0}$ has a limit iff for all $a < b$, $D(a, b) < \infty$.

(e) Show that $(x_n)_{n \geq 0}$ has a limit iff for all $a < b$, $(a, b) \in \mathbb{Q}^2$, $D(a, b) < \infty$.

Exercise 5. (Cauchy Criterion). Let $(x_n)_{n \geq 0} \subset \mathbb{R}$ be a sequence of (real numbers).

(a) Show that if $(x_n)_{n \geq 0}$ is Cauchy, then it has a unique accumulation point $\ell \in \mathbb{R}$ which is its limit.

(b) Show that if a sequence $(x_n)_{n \geq 0} \subset \mathbb{R}$ converges to $\ell \in \mathbb{R}$, then, it is Cauchy.

(c) Deduce the Cauchy criterion for sequences of real numbers.

SOLUTIONS

Exercise 1.

Question (1) : It is obvious that :

$$\inf_{p \geq n} x_p \leq x_n \leq \sup_{p \geq n} x_p,$$

since x_n is an element of $\{x_n, x_{n+1}, \dots\}$ on which we take the supremum or the infimum.

Question (2) : Let $y_n = \inf_{p \geq 0} x_p = \inf_{p \geq n} A_n$, where $A_n = \{x_n, x_{n+1}, \dots\}$ is a non-increasing sequence of sets : $\forall n \geq 0$,

$$A_{n+1} \subset A_n.$$

So the infimum on A_n increases. If y_n increases in $\overline{\mathbb{R}}$, its limit is its upper bound, finite or infinite. So

$$y_n \nearrow \underline{\lim} x_n,$$

is a finite or infinite number.

Question (3) : We also show that $z_n = \sup A_n$ decreases and $z_n \downarrow \overline{\lim} x_n$.

Question (4) : We recall that

$$-\sup \{x, x \in A\} = \inf \{-x, x \in A\}.$$

Which we write

$$-\sup A = \inf -A.$$

Thus,

$$-z_n = -\sup A_n = \inf -A_n = \inf \{-x_p, p \geq n\} ..$$

The right hand term tends to $-\overline{\lim} x_n$ and the left hand to $\underline{\lim} -x_n$ and so

$$-\overline{\lim} x_n = \underline{\lim} (-x_n).$$

Similarly, we show:

$$-\underline{\lim} (x_n) = \overline{\lim} (-x_n).$$

Question (5). These properties come from the formulas, where $A \subseteq \mathbb{R}, B \subseteq \mathbb{R}$:

$$\sup \{x + y, A \subseteq \mathbb{R}, B \subseteq \mathbb{R}\} \leq \sup A + \sup B.$$

In fact :

$$\forall x \in \mathbb{R}, x \leq \sup A$$

and

$$\forall y \in \mathbb{R}, y \leq \sup B.$$

Thus

$$x + y \leq \sup A + \sup B,$$

where

$$\sup_{x \in A, y \in B} x + y \leq \sup A + \sup B.$$

Similarly,

$$\inf(A + B) \geq \inf A + \inf B.$$

In fact :

$$\forall (x, y) \in A \times B, x \geq \inf A \text{ and } y \geq \inf B.$$

Thus

$$x + y \geq \inf A + \inf B.$$

Thus

$$\inf_{x \in A, y \in B} (x + y) \geq \inf A + \inf B$$

Application.

$$\sup_{p \geq n} (x_p + y_p) \leq \sup_{p \geq n} x_p + \sup_{p \geq n} y_p.$$

All these sequences are non-increasing. Taking infimum, we obtain the limits superior :

$$\overline{\lim} (x_n + y_n) \leq \overline{\lim} x_n + \overline{\lim} y_n.$$

Question (6) : Set

$$\underline{\lim} x_n = \overline{\lim} x_n,$$

Since :

$$\forall x \geq 1, y_n \leq x_n \leq z_n,$$

$$y_n \rightarrow \underline{\lim} x_n$$

and

$$z_n \rightarrow \overline{\lim} x_n,$$

we apply Sandwich Theorem to conclude that the limit of x_n exists and :

$$\lim x_n = \underline{\lim} x_n = \overline{\lim} x_n.$$

Exercise 2.

Question (a).

Thanks to question (4) of exercise 1, it suffices to show this property for one of the limits. Consider the limit superior and the three cases:

The case of a finite limit superior :

$$\underline{\lim} x_n = \ell \text{ finite.}$$

By definition,

$$z_n = \sup_{p \geq n} x_p \downarrow \ell.$$

So:

$$\forall \varepsilon > 0, \exists (N(\varepsilon) \geq 1), \forall p \geq N(\varepsilon), \ell - \varepsilon < x_p \leq \ell + \varepsilon.$$

Take less than that:

$$\forall \varepsilon > 0, \exists n_\varepsilon \geq 1 : \ell - \varepsilon < x_{n_\varepsilon} \leq \ell + \varepsilon.$$

We shall construct a subsequence converging to ℓ .

Let $\varepsilon = 1$:

$$\exists N_1 : \ell - 1 < x_{N_1} = \sup_{p \geq N_1} x_p \leq \ell + 1.$$

But if

$$(1.1) \quad z_{N_1} = \sup_{p \geq N_1} x_p > \ell - 1,$$

there surely exists an $n_1 \geq N_1$ such that

$$x_{n_1} > \ell - 1.$$

if not we would have

$$(\forall p \geq N_1, x_p \leq \ell - 1) \implies \sup \{x_p, p \geq N_1\} = z_{N_1} \leq \ell - 1,$$

which is contradictory with (1.1). So, there exists $n_1 \geq N_1$ such that

$$\ell - 1 < x_{n_1} \leq \sup_{p \geq N_1} x_p \leq \ell + 1.$$

i.e.

$$\ell - 1 < x_{n_1} \leq \ell + 1.$$

We move to step $\varepsilon = \frac{1}{2}$ and we consider the sequence $(z_n)_{n \geq n_1}$ whose limit remains ℓ . So, there exists $N_2 > n_1$:

$$\ell - \frac{1}{2} < z_{N_2} \leq \ell + \frac{1}{2}.$$

We deduce like previously that $n_2 \geq N_2$ such that

$$\ell - \frac{1}{2} < x_{n_2} \leq \ell + \frac{1}{2}$$

with $n_2 \geq N_1 > n_1$.

Next, we set $\varepsilon = 1/3$, there will exist $N_3 > n_2$ such that

$$\ell - \frac{1}{3} < z_{N_3} \leq \ell + \frac{1}{3}$$

and we could find an $n_3 \geq N_3$ such that

$$\ell - \frac{1}{3} < x_{n_3} \leq \ell + \frac{1}{3}.$$

Step by step, we deduce the existence of $x_{n_1}, x_{n_2}, x_{n_3}, \dots, x_{n_k}, \dots$ with $n_1 < n_2 < n_3 < \dots < n_k < n_{k+1} < \dots$ such that

$$\forall k \geq 1, \ell - \frac{1}{k} < x_{n_k} \leq \ell + \frac{1}{k},$$

i.e.

$$|\ell - x_{n_k}| \leq \frac{1}{k}.$$

Which will imply:

$$x_{n_k} \rightarrow \ell$$

Conclusion : $(x_{n_k})_{k \geq 1}$ is very well a subsequence since $n_k < n_{k+1}$ for all $k \geq 1$ and it converges to ℓ , which is then an accumulation point.

Case of the limit superior equal $+\infty$:

$$\overline{\lim} x_n = +\infty.$$

Since $z_n \uparrow +\infty$, we have : $\forall k \geq 1, \exists N_k \geq 1$,

$$z_{N_k} \geq k + 1.$$

For $k = 1$, let $z_{N_1} = \inf_{p \geq N_1} x_p \geq 1 + 1 = 2$. So there exists

$$n_1 \geq N_1$$

such that :

$$x_{n_1} \geq 1.$$

For $k = 2$: consider the sequence $(z_n)_{n \geq n_1+1}$. We find in the same manner

$$n_2 \geq n_1 + 1$$

and

$$x_{n_2} \geq 2.$$

Step by step, we find for all $k \geq 3$, an $n_k \geq n_{k-1} + 1$ such that

$$x_{n_k} \geq k.$$

Which leads to $x_{n_k} \rightarrow +\infty$ as $k \rightarrow +\infty$.

Case of the limit superior equal $-\infty$:

$$\overline{\lim} x_n = -\infty.$$

This implies : $\forall k \geq 1, \exists N_k \geq 1$, such that

$$z_{n_k} \leq -k.$$

For $k = 1, \exists n_1$ such that

$$z_{n_1} \leq -1.$$

But

$$x_{n_1} \leq z_{n_1} \leq -1$$

Let $k = 2$. Consider $(z_n)_{n \geq n_1+1} \downarrow -\infty$. There will exist $n_2 \geq n_1 + 1$:

$$x_{n_2} \leq z_{n_2} \leq -2$$

Step by step, we find $n_{k+1} < n_k$ in such a way that $x_{n_k} < -k$ for all k bigger than 1. So

$$x_{n_k} \rightarrow +\infty$$

Question (b).

Let ℓ be an accumulation point of $(x_n)_{n \geq 1}$, the limit of one of its subsequences $(x_{n_k})_{k \geq 1}$. We have

$$y_{n_k} = \inf_{p \geq n_k} x_p \leq x_{n_k} \leq \sup_{p \geq n_k} x_p = z_{n_k}$$

The left hand side term is a subsequence of (y_n) tending to the limit inferior and the right hand side is a subsequence of (z_n) tending to the limit superior. So we will have:

$$\underline{\lim} x_n \leq \ell \leq \overline{\lim} x_n,$$

which shows that $\underline{\lim} x_n$ is the smallest accumulation point and $\overline{\lim} x_n$ is the largest.

Question (c). If the sequence $(x_n)_{n \geq 1}$ has a limit ℓ , it is the limit of all its subsequences, so subsequences tending to the limits superior and inferior. Which answers question (b).

Question (d). We answer this question by combining point (d) of this exercise and point (6) of the exercise 1.

Exercise 3. Let $(x_n)_{n \geq 0}$ be a non-decreasing sequence, we have:

$$z_n = \sup_{p \geq n} x_p = \sup_{p \geq 0} x_p, \forall n \geq 0.$$

Why? Because by increasingness,

$$\{x_p, p \geq 0\} = \{x_p, 0 \leq p \leq n-1\} \cup \{x_p, p \geq n\}$$

Since all the elements of $\{x_p, 0 \leq p \leq n-1\}$ are smaller than that of $\{x_p, p \geq n\}$, the supremum is achieved on $\{x_p, p \geq n\}$ and so

$$\ell = \sup_{p \geq 0} x_p = \sup_{p \geq n} x_p = z_n$$

Thus

$$z_n = \ell \rightarrow \ell.$$

We also have $y_n = \inf \{x_p, 0 \leq p \leq n\} = x_n$ which is a non-decreasing sequence and so converges to $\ell = \sup_{p \geq 0} x_p$.

Exercise 4.

Let $\ell \in \overline{\mathbb{R}}$ having the indicated property. Let ℓ' be a given accumulation point.

$$(x_{n_k})_{k \geq 1} \subseteq (x_n)_{n \geq 0} \text{ such that } x_{n_k} \rightarrow \ell'.$$

By hypothesis this subsequence (x_{n_k}) has in turn a subsubsequence $(x_{n_{k(p)}})_{p \geq 1}$ such that $x_{n_{k(p)}} \rightarrow \ell$ as $p \rightarrow +\infty$.

But as a subsequence of (x_{n_k}) ,

$$x_{n_{k(\ell)}} \rightarrow \ell'.$$

Thus

$$\ell = \ell'.$$

Applying that to the limit superior and limit inferior, we have:

$$\overline{\lim} x_n = \underline{\lim} x_n = \ell.$$

And so $\lim x_n$ exists and equals ℓ .

Exercise 5.

Question (a). If ν_{2k} finite and ν_{2k+1} infinite, it then has exactly k up-crossings : $[x_{\nu_{2j-1}}, x_{\nu_{2j}}], j = 1, \dots, k : D(a, b) = k$.

Question (b). If ν_{2k+1} finite and ν_{2k+2} infinite, it then has exactly k up-crossings: $[x_{\nu_{2j-1}}, x_{\nu_{2j}}], j = 1, \dots, k : D(a, b) = k$.

Question (c). If all the ν'_j s are finite, then, there are an infinite number of up-crossings : $[x_{\nu_{2j-1}}, x_{\nu_{2j}}], j \geq 1k : D(a, b) = +\infty$.

Question (d). Suppose that there exist $a < b$ rationals such that $D(a, b) = +\infty$. Then all the ν'_j s are finite. The subsequence $x_{\nu_{2j-1}}$ is strictly below a . So its limit inferior is below a . This limit inferior is an accumulation point of the sequence $(x_n)_{n \geq 1}$, so is more than $\underline{\lim} x_n$, which is below a .

Similarly, the subsequence $x_{\nu_{2j}}$ is strictly below b . So the limit superior is above a . This limit superior is an accumulation point of the sequence $(x_n)_{n \geq 1}$, so it is below $\overline{\lim} x_n$, which is directly above b . Which leads to:

$$\underline{\lim} x_n \leq a < b \leq \overline{\lim} x_n.$$

That implies that the limit of (x_n) does not exist. In contrary, we just proved that the limit of (x_n) exists, meanwhile for all the real numbers a and b such that $a < b$, $D(a, b)$ is finite.

Now, suppose that the limit of (x_n) does not exist. Then,

$$\underline{\lim} x_n < \overline{\lim} x_n.$$

We can then find two rationals a and b such that $a < b$ and a number ϵ such that $0 < \epsilon$, all the

$$\underline{\lim} x_n < a - \epsilon < a < b < b + \epsilon < \overline{\lim} x_n.$$

If $\underline{\lim} x_n < a - \epsilon$, we can return to question (a) of exercise 2 and construct a subsequence of (x_n) which tends to $\underline{\lim} x_n$ while remaining below $a - \epsilon$. Similarly, if $b + \epsilon < \overline{\lim} x_n$, we can create a subsequence of (x_n) which tends to $\overline{\lim} x_n$ while staying above $b + \epsilon$. It is evident with these two sequences that we could define with these two sequences all

ν_j finite and so $D(a, b) = +\infty$.

We have just shown by contradiction that if all the $D(a, b)$ are finite for all rationals a and b such that $a < b$, then, the limit of $(x)_n$ exists.

Exercise 5. Cauchy criterion in \mathbb{R} .

Suppose that the sequence is Cauchy, *i.e.*,

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_p - x_q) = 0.$$

Then let $x_{n_{k,1}}$ and $x_{n_{k,2}}$ be two subsequences converging respectively to $\ell_1 = \underline{\lim} x_n$ and $\ell_2 = \overline{\lim} x_n$. So

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_{n_{p,1}} - x_{n_{q,2}}) = 0.$$

, By first letting $p \rightarrow +\infty$, we have

$$\lim_{q \rightarrow +\infty} \ell_1 - x_{n_{q,2}} = 0.$$

Which shows that ℓ_1 is finite, else $\ell_1 - x_{n_{q,2}}$ would remain infinite and would not tend to 0. By interchanging the roles of p and q , we also have that ℓ_2 is finite.

Finally, by letting $q \rightarrow +\infty$, in the last equation, we obtain

$$\ell_1 = \underline{\lim} x_n = \overline{\lim} x_n = \ell_2.$$

which proves the existence of the finite limit of the sequence (x_n) .

Now suppose that the finite limit ℓ of (x_n) exists. Then

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_p - x_q) = \ell - \ell = 0.$$

0 Which shows that the sequence is Cauchy.

2. Miscelleanuous facts

FACT 1. For any $a \in \mathbb{R}$,

$$|e^{ia} - 1| = \sqrt{2(1 - \cos a)} \leq 2 |\sin(a/2)| \leq 2 |a/2|^\delta.$$

This is easy for $|a/2| > 1$. Indeed for $\delta > 0$, $|a/2|^\delta > 0$ and

$$2 |\sin(a/2)| \leq 2 \leq 2 |a/2|^\delta$$

Now for $|a/2| > 1$, we have the expansion

$$\begin{aligned} 2(1 - \cos a) &= a^2 - \sum_{k=2}^{\infty} (-1)^2 \frac{a^{2k}}{(2k)!} = x^2 - 2 \sum_{k \geq 2, k \text{ even}}^{\infty} \frac{a^{2k}}{(2k)!} - \frac{a^{2(k+1)}}{(2(k+1))!} \\ &= a^2 - 2x^{2(k+1)} \sum_{k \geq 2, k \text{ even}}^{\infty} \frac{1}{(2k)!} \left\{ \frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\}. \end{aligned}$$

For each $k \geq 2$, for $|a/2| < 1$,

$$\left\{ \frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\} \geq \left\{ \frac{1}{4} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\} \geq 0.$$

Hence

$$2(1 - \cos a) \leq a^2.$$

But for $|a/2|$, the function $\delta \mapsto |a/2|^\delta$ is non-increasing δ , $0 \leq \delta \leq 1$.

Then

$$\sqrt{2(1 - \cos a)} \leq |a| = 2 |a/2|^1 \leq 2 |a/2|^\delta.$$

Bibliography

- [1] Bauer, H.(1981). *Probability Theory and Elements of Measure Theory*. Holt, Rinehart, and Winston, New-York.
- [2] Billingsley, P.(1968). *Convergence of Probability measures*. John Wiley, New-York.
- [3] Dudley, R. M.(1989). Real Analysis and probability. Wadsworth, Pacific Grove.
- [4] Gutt, A.(2005). *Probability : A Graduate Course*. Springer-Verlag.
- [5] Dudley, R. M.(1989). Real Analysis and probability. Wadsworth, Pacific Grove.
- [6] Lo, G.S.(2016). A Course on Elementary Probability Theory. SPAS Editions. Saint-Louis, Calgary, Abuja. Doi : 10.16929/sbs/2016.0003
- [7] Lo, G.S.(2016). Cours Elementaire de Théorie de Probabilités. SPAS Editions. Saint-Louis, Calgary, Abuja. Doi : 10.16929/sbs/2016.0004
- [8] Lo, G.S.(2016). *Introduction to stochastic processes*. Spas Textbooks Series.
- [9] Lo, G.S.(2016). *Mathematical Foundation to Probability Theory*. Spas Textbooks Series.
- [10] Loève, Michel.(1997). *Probability Theory I*. Springer-Verlag, 4th Edition.
- [11] A. W. van der Vaart and J. A. Wellner(1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New-York.
- [12] van der Vaart, A.W. Asymptotics Statistics. (2000). *Cambridge*